What is Logical Validity?

Whatever other merits proof-theoretic and model-theoretic accounts of validity may have, they are not remotely plausible as accounts of the meaning of 'valid'. And not just because they involve technical notions like 'model' and 'proof' that needn't be possessed by a speaker who understands the concept of valid inference. The more important reason is that competent speakers may agree on the model-theoretic and proof-theoretic facts, and yet disagree about what's valid.

Consider for instance the usual model-theoretic account of validity for sentential logic: an argument is sententially valid iff any total function from the sentences of the language to the values T,F that obeys the usual compositional rules and assigns T to all the premises of the argument also assigns T to the conclusion. Let's dub this *classical sentential validity*.[1] There's no doubt that this is a useful notion, but it couldn't possibly be what we *mean* by 'valid' (or even by 'sententially valid', i.e. 'valid by virtue of sentential form'). The reason is that even those who reject classical sentential logic will agree that the sentential inferences that the classical logician accepts are valid in *this* sense. For instance, someone who thinks that "disjunctive syllogism" (the inference from $A \vee B$ and $\neg A$ to B) is not a valid form of inference will, if she accepts a bare minimum of mathematics,[2] agree that the inference is *classically* valid, and will say that that just shows that classical validity outruns genuine *validity*. Those who accept disjunctive syllogism don't just believe it *classically* valid, which is beyond serious contention; they believe it *valid*.

This point is in no way peculiar to classical logic. Suppose an advocate of a sentential

[1] Note that this is very different from saying that validity consists in necessarily preserving truth; that account will be considered in Section 1. The model-theoretic account differs from the necessary truth-preservation account in being purely mathematical: it invokes functions that assign the object T or F to each sentence (and that obey the compositional rules of classical semantics), without any commitment to any claims about how T and F relate to truth and falsity. For instance, it involves no commitment to the claim that each sentence is either true or false and not both, or that the classical compositional rules *as applied to truth and falsity* are correct.

[2] And there's no difficulty in supposing that the non-classical logician does so, or even, that she accepts classical mathematics across the board: she may take mathematical objects to obey special non-logical assumptions that make classical reasoning "effectively valid" within mathematics.

logic without disjunctive syllogism offers a model theory for her logic—e.g. one on which an argument is sententially valid iff any assignment of one of the values T, U, F to the sentences of the language that obeys certain rules and gives the premises a value other than F also gives the conclusion other than F ("*LP-validity*"). This may make only her preferred sentential inferences come out "valid", but it would be subject to a similar objection if offered as an account of the meaning of 'valid': classical logicians who accept more sentential inferences, and other non-classical logicians who accept fewer, will agree with her as to what inferences meet this definition, but will disagree about which ones are valid. Whatever logic L one advocates, one should recognize a distinction between the concept 'valid-in-L' and the concept 'valid'.[3]

The same point holds (perhaps even more obviously) for provability in a given deductive system: even after we're clear that a claim does or doesn't follow from a given deductive system for sentential logic, we can disagree about whether it's valid.

I don't want to make a big deal about definition or meaning: the point I'm making can be made in another way. It's that advocates of different logics presumably disagree about something—and something more than just how to use the term 'valid', if their disagreement is more than verbal. It would be nice to know what it is they disagree about. And they don't disagree about what's classically valid (as defined either model-theoretically or proof-theoretically); nor about what's intuitionistically valid, or LP-valid, or whatever. So *what do they disagree about?* That is the main topic of the paper, and will be discussed in Sections 1-4.

Obviously model-theoretic and proof-theoretic accounts of validity are important. So another philosophical issue is to explain what their importance is, given that it is not to explain the concept of validity. Of course one obvious point can be made immediately: the model

[3] I'm tempted to call my argument here a version of Moore's Open Question Argument: a competent speaker may say "Sure, that inference is *classically* valid, but is it *valid*?" (or, "Sure, that inference is *LP-invalid*, but is it *invalid*?"). But only on a sympathetic interpretation of Moore, in which he isn't attempting a general argument against there being an acceptable naturalistic definition but rather is simply giving a way to elicit the implausibility of particular naturalistic definitions. In the next section I will consider a proposal for a naturalistic definition of 'valid' which (though I oppose) I do not take to be subject to the kind of "open question" argument I'm employing here.

theories and proof theories for classical logic, LP, etc. are effective tools for ascertaining what is and isn't classically valid, LP-valid, etc.; so to someone convinced that one of these notions extensionally coincides with genuine validity, the proof-theory and model-theory provide effective tools for finding out about validity. But there's much more than this obvious point to be said about the importance of model-theoretic and proof-theoretic accounts; that will be the topic of Sections 5 and 6.

1: Necessarily preserving truth.

One way to try to explain the concept of validity is to define it in other (more familiar or more basic) terms. As we've seen, any attempt to use model theory or proof theory for this purpose would be hopeless; but there is a prominent alternative way of trying to define it. In its simplest form, validity is explained by saying that an inference (or argument)[4] is valid iff it preserves truth by logical necessity.

It should be admitted at the start that there are non-classical logics (e.g. some relevance logics, dynamic logics, linear logic) whose point seems to be to require more of validity than logically necessary preservation of truth. Advocates of these logics may want their inferences to necessarily preserve truth, but they want them to do other things as well: e.g. to preserve conversational relevance, or what's settled in a conversation, or resource use, and so forth. There are other logics (e.g. intuitionist logic) whose advocates may or may not have such additional goals. Some who advocate intuitionistic logic (e.g. Dummett) think that reasoning classically leads to error; which perhaps we can construe as, possibly fails to preserve truth. But others use intuitionistic logic simply in order to get proofs that are more informative than classical, because constructive; insofar as *those* intuitionists reserve 'valid' for intuitionistic validity, they too are imposing additional goals of quite a different sort than truth preservation.

---

[4] The term 'inference' can mislead: as Harman has pointed out many times, inferring is naturally taken as a dynamic process in which one comes to new beliefs, and inference in that sense is not directly assessable in terms of validity. But there is no obvious non-cumbersome term for what it is that's valid that would be better—'argument' has the same problem as 'inference', and other problems besides. (A cumbersome term for what's valid is "pair $<\Gamma, B>$ where *B* is a formula and Γ a set of formulas".)

While it is correct that there are logicians for whom truth preservation is far from the sole goal, this isn't of great importance for my purposes. That's because my interest is with *what people who disagree in logic are disagreeing about*; and if proponents of one logic want that logic to meet additional goals that proponents of another logic aren't trying to meet, and reject inferences that the other logic accepts only because of the difference of goals, then the apparent disagreement in logic seems merely verbal.

I take it that logically necessary truth preservation is a good first stab at what advocates of classical logic take logic to be concerned with. My interest is with those who share the goals of the classical logician, but who are in non-verbal disagreement as to which inferences are valid. This probably doesn't include any advocates of dynamic logics or linear logic, but it includes some advocates of intuitionist logic and quantum logic, and most advocates of various logics designed to cope with vagueness and/or the semantic paradoxes. So these will be my focus. The claim at issue in this section is that *genuine* logical disagreement is disagreement about which inferences preserve truth by logical necessity.

Having set aside linear logic and the like, a natural reaction to the definition of validity as preservation of truth by logical necessity is that it isn't very informative: logical necessity looks awfully close to validity, indeed, logically necessary truth is just the special case of validity for 0-premise arguments. One can make the account slightly more informative by explaining logical necessity in terms of some more general notion of necessity together with some notion of logical form, yielding that an argument is valid iff (necessarily?) *every argument that shares its logical form* necessarily preserves truth.[5] Even so, it could well be worried that the use of the notion of

[5] The idea of one inference "sharing the logical form of" another requires clarification. It is easy enough to explain 'shares the *sentential* form of', 'shares the *quantificational* form of', and so on, but explaining 'shares the logical form of' is more problematic, since it depends on a rather open-ended notion of what aspects of form are logical. But perhaps, in the spirit of the remarks in Tarski 1936 on there being no privileged notion of logical constant, we should say that we don't really need a general notion of validity, but only a notion of validity relative to a given choice of which terms count as logical constants (so that validity then subdivides into sentential validity, first order quantificational validity, first order quantificational-plus-identity validity, and so on). And in explaining e.g. sentential validity in the manner contemplated, we need no more than the idea of sharing sentential form.

necessity is helping ourselves to something that ought to be explained.

This worry becomes especially acute when we look at the way that logical necessity needs to be understood for the definition of validity in terms of it to get off the ground. Consider logics according to which excluded middle is not valid. Virtually no such logic accepts of any instance of excluded middle that it is not true: that would seem tantamount to accepting a sentence of form $\neg(B \vee \neg B)$, which in almost any logic requires accepting $\neg B$, which in turn in almost any logic requires accepting $B \vee \neg B$ and hence is incompatible with the rejection of this instance of excluded middle. To say that $B \vee \neg B$ is not *necessarily* true would seem to raise a similar problem: it would seem to imply that it is *possibly* not true, which would seem to imply that there's a possible state of affairs in which $\neg(B \vee \neg B)$; but then, by the same argument, that would be a possible state of affairs in which $\neg B$ and hence $B \vee \neg B$, and we are again in contradiction. Given this, how is one who regards some instances of excluded middle as invalid to maintain the equation of validity with logically necessary truth? The only obvious way is to resist the move from 'it isn't logically necessary that *p*' to 'there's a possible state of affairs in which $\neg p$'. I think we must do that; but if we do, I think we remove any sense that we were dealing with a sense of necessity that we have a grasp of independent of the notion of logical truth.[6]

But let's put aside any worry that the use of necessity in explaining validity is helping ourselves to something that ought to be explained. I want to object to the proposed definition of validity in a different way: that it simply gives the wrong results about what's valid. That is: it gives results that are at variance with our ordinary notion of validity. Obviously it's possible to simply insist that by 'valid' one will simply mean 'preserves truth by logical necessity'. But as

[6] An alternative move would be to say that one who rejects excluded middle doesn't believe that excluded middle isn't valid, but either (i) merely fails to believe that it is, or (ii) *rejects* that it is, without believing that it isn't. But (i) seems clearly wrong: it fails to distinguish the advocate of a logic that rejects excluded middle with someone agnostic between that logic and classical. (ii) is more defensible, though I don't think it's right. (While I think an opponent of excluded middle does reject some instances of excluded middle while not believing their negation, I think that he or she regards excluded middle as not valid.) But in any case, the distinction between rejection and belief in the negation is not typically recognized by advocates of the necessary truth-preservation account validity, and brings in ideas that suggest the quite different account of validity to be advanced in Section 2.

we'll see, this definition would have surprising and unappealing consequences, which I think should dissuade us from using 'valid' in this way.

Let $A_1,...,A_n \Rightarrow B$ mean that the argument from $A_1, ..., A_n$ to B is valid.[7] The special case $\Rightarrow B$ (that the argument from no premises to B is valid) means in effect that B is a valid *sentence*, i.e. is in some sense logically necessary. The proposed definition of valid argument tries to explain

(I) $\qquad A_1,...,A_n \Rightarrow B$

as

(II_T) $\qquad \Rightarrow \text{True}(\langle A_1\rangle) \wedge ... \wedge \text{True}(\langle A_n\rangle) \rightarrow \text{True}(\langle B\rangle)$.

This is an attempt to explain validity of inferences in terms of the validity (logical necessity) of single sentences. I think that any attempt to do this is bound to fail.

The plausibility of thinking that (I) is equivalent to (II_T) depends, I think, on two purported equivalences: first, between (I) and

(II) $\qquad \Rightarrow A_1 \wedge ... \wedge A_n \rightarrow B$;

second, between (II) and (II_T).[8]

An initial point to make about this is that while (I) is indeed equivalent to (II) in classical and intuitionist logic, there are many nonclassical logics in which it is not. (These include even supervaluational logic, which is sometimes regarded as classical.) In *most* standard logics, (II)

---

[7] I take the A_is and B in '$A_1,...,A_n \Rightarrow B$' to be variables ranging over sentences, and the '$\Rightarrow$' to be a predicate. (So in a formula such as $\Rightarrow A \rightarrow B$, what comes after the '$\Rightarrow$' should be understood as a complex singular term with two free variables, in which the '$\rightarrow$' is a function symbol. A claim such as $A \Rightarrow \text{True}(\langle A\rangle)$ should be understood as saying that (for each sentence A) the inference from A to the result of predicating 'True' of the structural name of A is valid.) In other contexts I'll occasionally use italicized capital letters as abbreviations of formulas or as schematic letters for formulas; I don't think any confusion is likely to result.

[8] Or alternatively, first between (I) and

(I_T) $\qquad \text{True}(\langle A_1\rangle), ..., \text{True}(\langle A_n\rangle) \Rightarrow \text{True}(\langle B\rangle)$;

second, between (I_T) and (II_T). The latter purported equivalence is of course a special case of the purported equivalence between (I) and (II), so the discussion in the text still applies.

requires (I). But there are many logics in which conditional proof fails, so that (I) does not require (II). (Logics where ∧-Elimination fails have the same result.) In such logics, we wouldn't expect (I) to require (II_T), so validity would not require logically necessary truth preservation.

Perhaps this will seem a quibble, since *many* of those who reject conditional proof want to introduce a notion of "super-truth" or "super-determinate truth", and will regard (I) as equivalent to

(II_{ST}) $\Rightarrow$ Super-true($\langle A_1 \rangle$) $\wedge$ ... $\wedge$ Super-true($\langle A_n \rangle$) $\rightarrow$ Super-true($\langle B \rangle$).

In that case, they are still reducing the validity of an inference to the validity of a conditional, just a different conditional, and we would have a definitional account of validity very much in the spirit of the first. I will be arguing, though, that the introduction of super-truth doesn't help: (I) not only isn't equivalent to (II_T), it isn't equivalent to (II_{ST}) either, whatever the notion of super-truth. Validity isn't the preservation of either truth or "super-truth" by logical necessity.

To evaluate the proposed reduction of validity to preservation of truth or super-truth by logical necessity, we need to first see how well validity so defined coincides in extension with validity as normally understood. Here there's good news and bad news. The good news is that (at least insofar as vagueness can be ignored, as I will do) there is *very* close agreement; the bad news is that where there is disagreement, the definition in terms of logically necessary preservation of truth (or super-truth) gives results that seem highly counterintuitive.

The good news is implicit in what I've already said, but let me spell it out. Presumably for at least a wide range of sentences A_1, ..., A_n and B, claim (II) above is equivalent to (II_T), and claim (I) is equivalent to the (I_T) of note 8. (I myself think these equivalences holds for *all* sentences, but I don't want to presuppose controversial views. Let's say that (II) is equivalent to (II_T) (and (I) to (I_T)) *at least* for all "ordinary" sentences A_1, ..., A_n, B, leaving unspecified where exceptions might lie if there are any.) And presumably when A_1, ..., A_n and B are "ordinary", (I) is equivalent to (II) (and (I_T) to (II_T)).[9] In that case, we have

[9] This actually may not be so in the presence of vagueness: in many logics of vagueness (e.g. Lukasiewicz logics), (I) can hold when (II) doesn't. Admittedly, in Lukasiewicz logics there is a notion of super-determinate truth for which (I) does correspond to something close to (II), viz. (II_{ST}). But as I've argued elsewhere (e.g. Field 2008, Ch. 5), the presence of such a notion of

(GoodNews) The equivalence of (I) to (II_T) holds at least for "ordinary" sentences: for those, validity does coincide with preservation of truth by logical necessity.

(Presumably those with a concept of super-truth think that for sufficiently "ordinary" sentences it coincides with truth; if so, then the good news also tells us that (I) coincides with the logically necessary preservation of *super-truth*.)

Despite this good news for the attempt to define validity in terms of logically necessary truth preservation, the bad news is that the equivalence of (I) to either (II_T) or (II_{ST}) can't plausibly be maintained for *all* sentences. The reason is that in certain contexts, most clearly the semantic paradoxes but possibly for vagueness too, this account of validity requires a wholly implausible divorce between which inferences are declared valid and which ones are deemed acceptable to use in reasoning (even static reasoning, for instance in determining reflective equilibrium in one's beliefs). In some instances, the account of validity would require having to reject the validity of logical reasoning that one finds completely acceptable and important. In other instances, it would even require declaring reasoning that one thinks leads to error to be nonetheless valid!

I can't give a complete discussion here, because the details will depend on how one deals either with vagueness or with the "non-ordinary" sentences that arise in the semantic paradoxes. I'll focus on the paradoxes, where I'll sketch what I take to be the two most popular solutions and show that in the context of each of them, the proposed definition of validity leads to very bizarre consequences. (Of course the paradoxes themselves force some surprising consequences, but the bizarre consequences of the proposal for validity go way beyond that.)

► **Illustration 1:** It is easy to construct a "Curry sentence" *K* that is equivalent (given uncontroversial assumptions) to "If True(⟨*K*⟩) then 0=1". This leads to an apparent paradox. The most familiar reasoning to the paradox first argues from the assumption that True(⟨*K*⟩) to the conclusion that 0=1, then uses conditional proof to infer that *if True(⟨K⟩) then 0=1*, then argues from that to the conclusion that True(⟨*K*⟩); from which we then repeat the original reasoning to '0=1', but this time with True(⟨*K*⟩) as a previously established result rather than as an assumption. Many theories of truth (this includes most supervaluational theories and revision theories as well as most non-classical theories) take the sole problem with this reasoning to be its use of conditional proof. In particular, they agree that the reasoning from the

---

super-determinate truth in these logics is a crippling defect: it spoils them as logics of vagueness. In an adequate non-classical logic of vagueness, (I) won't be equivalent to anything close to (II).

assumption of 'True(⟨K⟩)' to '0=1' is perfectly acceptable (given the equivalence of *K* to "If True(⟨*K*⟩) then 0=1"), and that the reasoning from "If True(⟨*K*⟩) then 0=1" to *K* and from that to "True(⟨*K*⟩)" is acceptable as well. I myself think that the best solutions to the semantic paradoxes take this position on the Curry paradox.

But what happens if we accept such a solution, but define 'valid' in a way that requires truth-preservation? In that case, though we can legitimately reason from *K* to '0=1' (via the intermediate 'True(⟨*K*⟩)', we can't declare the inference "valid". For to say that it is "valid" in this sense is to say that True(⟨*K*⟩) → True(⟨0=1⟩), which yields True(⟨*K*⟩) → 0=1, which is just *K*; and so calling the inference "valid" in the sense defined would lead to absurdity. That's very odd: this theorist accepts the reasoning from *K* to 0=1 as completely legitimate, and indeed *it's only because he reasons in that way that he sees that he can't accept K*; and yet on the proposed definition of 'valid' he is precluded from calling that reasoning "valid". ◂

▸ **Illustration 2:** Another popular resolution of the semantic paradoxes (the truth-value gap resolution) has it that conditional proof is fine, but it isn't always correct to reason from *A* to True(⟨*A*⟩). Many people who hold this (those who advocute "Kleene-style gaps") do think you can reason from True(⟨*A*⟩) to True(⟨True(⟨*A*⟩)⟩); and so, by conditional proof, they think you should accept the conditional True(⟨*A*⟩)→True(⟨True(⟨*A*⟩)⟩). Faced with a Curry sentence, or even a simpler Liar sentence *L*, their claim is that *L* isn't true, and that the sentence *⟨L⟩ isn't true* (which is equivalent to *L*) isn't true either. There is an obvious oddity in such resolutions of the paradoxes: in claiming that one should believe *L* but not believe it true, the resolution has it that truth isn't the proper object of belief.[10] But odd or not, this sort of resolution of the paradoxes is quite popular.

But the advocate of such a theory *who goes on to define "valid" in terms of necessary preservation of truth* is in a far odder situation. First, this theorist accepts the reasoning to the conclusion ¬*True(⟨L⟩)*—he regards ¬*True(⟨L⟩)* as essentially a theorem. But since he regards the conclusion as not true, then he regards the (0-premise) reasoning to it as "invalid", on the definition in question: he accepts a conclusion on the basis of reasoning, while declaring that reasoning "invalid". This is making an already counterintuitive theory sound even worse, by a perverse definition of validity.

But wait, there's more! Since the view accepts both ¬True(⟨*L*⟩) and ¬True(⟨¬True(⟨*L*⟩)⟩), and doesn't accept contradictions, it obviously doesn't accept the reasoning from ¬True(⟨*L*⟩) to True(⟨¬True(⟨*L*⟩)⟩) as good. But on the proposed definition of "valid", the view does accept it as valid! For on the proposed definition, that simply means that True(⟨¬True(⟨*L*⟩)⟩) → True(⟨True(⟨¬True(⟨*L*⟩)⟩)⟩), and as remarked at the start, the view does accept all claims of form True(⟨*A*⟩)→True(⟨True(⟨*A*⟩)⟩). On the definition of validity, not only can good logical reasoning come out

[10] Others, even some who accept the resolution of the truth paradoxes in Illustration 1, introduce a notion of Supertruth on which it isn't always correct to reason from *A* to Supertrue(⟨*A*⟩) but is correct to assert Supertrue(⟨*A*⟩)→Supertrue(⟨Supertrue(⟨*A*⟩)⟩) This resolution leads to a Liar-like sentence *L**, and asserts that *L** isn't supertrue and that '⟨*L**⟩ isn't supertrue' isn't supertrue either. This may be slightly less odd, since it says only that the technical notion of supertruth isn't the proper object of belief.

invalid, but fallacious reasoning can come out valid. ◂

These problems for defining validity in terms of necessary preservation of truth are equally problems for defining validity in terms of necessary preservation of supertruth: we need only consider paradoxes of supertruth constructed in analogy to the paradoxes of truth (e.g. a modified Curry sentence that asserts that if it's supertrue then 0=1, and a modified Liar that asserts that it isn't supertrue). Then given any resolution of such paradoxes, we reason as before to show the divorce between the super-truth definition of validity and acceptable reasoning.

I've said that as long as we put vagueness aside (see note 9), it's only for fairly "non-ordinary" inferences that the definition of validity as preservation of truth (or supertruth) by logical necessity is counterintuitive: for most inferences that don't crucially involve vague terms, the definition gives extremely natural results. But that is because it is only for such non-ordinary inferences that the approach leads to different results than the approach I'm about to recommend!

In the next section I'll recommend a different approach to validity, whose central idea is that validity attributions *regulate our beliefs*. Considerations about whether an inference preserves truth are certainly highly *relevant* to the regulation of belief. Indeed, on my own approach to the paradoxes and some others, the following all hold:[11]

> (A) Logically necessary truth-preservation *suffices for* validity in the regulative sense; e.g., if $\Rightarrow \text{True}(\langle A\rangle) \rightarrow \text{True}(\langle B\rangle)$ then one's degree of belief in B should be at least that of A.
>
> (B) Logically necessary truth is *necessary for* the validity of *sentences*: it is only for inferences with at least one premise that the implication from validity to truth-preservation fails.

(C) If an argument is valid, there can be no *clear* case of its failing to preserve truth.

[11] (A) follows from the equivalence between $\text{True}(\langle C\rangle)$ and C for arbitrary C, modus ponens, and the principle (VP) to be given in Section 2. (B) holds since it's only for inferences with at least one premise that conditional proof is relevant. (C) and (D) follow from what I've called "restricted truth preservation", e.g. in Field 2008 p. 148: on a theory like mine, valid arguments with unproblematically true premises preserve truth.

(D) If an argument is valid, then we should believe that it is truth-preserving to at least as high a degree as we believe the conjunction of its premises.

This collection of claims seems to me to get at what's right in truth-preservation definitions of validity, without the counterintuitive consequences.

2. Validity and the regulation of belief.

The necessary truth-preservation approach to explaining the concept of validity tried to define that concept in other (more familiar or more basic) terms. I'll briefly mention another approach that takes this form, in Section 3; but first I'll expound what I think a better approach, which is to leave 'valid' undefined but to give an account of its "conceptual role". That's how we explain negation, conjunction, etc.; why not 'valid' too?

The basic idea for the conceptual role is

$(VB)_a$ To regard an inference or argument as valid is (in large part anyway) to accept a constraint on belief: one that prohibits fully believing its premises without fully believing its conclusion.[12] (At least for now, let's add that the prohibition should be "due to logical form": for any other argument of that form, the constraint should also prohibit fully believing the premises without fully believing the conclusion.[13] This addition may no longer be needed once we move to the expanded version in Section 2(d).)

The underlying idea here is that *a disagreement about validity (insofar as it isn't merely verbal) is a disagreement about what constraints to impose on one's belief system.*

It would be natural to rewrite this principle as saying that to regard an inference as valid is to hold that one *shouldn't* fully believing its premises without fully believing its conclusion. And it's then natural to go from the rewritten principle about what *we regard as* valid to the following principle about what *is* valid:

---

[12] Note that the constraint on belief in $(VB)_a$ is a static constraint, not a dynamic one: it doesn't dictate that if a person discovers new consequences of his beliefs he should believe those consequences (rather than abandoning some of his current beliefs).

[13] The remarks on logical form in note 5 then apply here too.

$(VB)_n$ If an argument is valid, then we shouldn't fully believe the premises without fully believing the conclusion.

(The subscripts on $(VB)_a$ and $(VB)_n$ stand for 'attitudinal' and 'normative'.) I'll play up the differences between $(VB)_a$ and $(VB)_n$ in Section 4, and explain why I want to take a formulation in the style of $(VB)_a$ as basic. But in the rest of the paper the difference will play little role; so until then I'll work mainly with the simpler formulation $(VB)_n$. And since the distinction won't matter until then, I'll usually just leave the subscript off.

In either form, (VB) needs both qualification and expansion. One way in which it should be qualified is an analog of the qualification already made for necessary truth-preservation accounts: (VB) isn't intended to apply to logics (such as linear logic) whose validity relation is designed to reflect matters such as resource use that go beyond what the classical logician is concerned with; restricting what counts as "valid" merely because of such extra demands on validity isn't in any non-verbal sense a disagreement with classical logic. This is a point on which the legitimacy of belief approach to validity and the necessary truth-preservation approach are united; they diverge only on whether the core concern of classical logic (and many non-classical logics too, though not linear logic) is to be characterized in terms of legitimacy of belief or necessary truth-preservation.

The other qualifications of (VB) mostly concern (i) the computational complexity of logic and (ii) the possibility of logical disagreement among informed agents. The need for such qualifications is especially clear when evaluating other people. To illustrate (i), suppose that I have after laborious effort proved a certain unobvious mathematical claim, by a proof formalizable in standard set theory, but that you don't know of the proof; then (especially if the claim is one that seems prima facie implausible),[14] there seems a clear sense in which I think you should *not* believe the claim without proof, even though you believe the standard set-theoretic axioms; i.e. you should violate my prohibition. To illustrate (ii), suppose that you and I have long accepted restrictions on excluded middle to handle the semantic paradoxes, and that you have developed rather compelling arguments that this is the best way to go; but suppose that I have recently found new considerations,

---

[14] E.g. the existence of a continuous function taking the unit interval onto the unit square, or the possibility of decomposing a sphere into finitely many pieces and rearranging them to get two spheres each of the same size as the first.

not known to you, for rejecting these arguments and for insisting on one of the treatments of the paradoxes within classical logic. On this scenario, I take excluded middle to be valid; but there seems a clear sense in which I think you *shouldn't* accept arguments which turn on applying excluded middle in a way that isn't licensed *by your theory*, which again means violating my prohibition.

To handle such examples, one way to go would be to suppose that to regard an inference as valid is to accept the above constraint on belief only as applied to those who recognize it as valid. But that is awkward in various ways. A better way to go (as John MacFarlane convinced me a few years back)[15] is to say that we recognize multiple constraints on belief, which operate on different levels and may be impossible to simultaneously satisfy. When we are convinced that a certain proof from premises is valid, we think that in some "non-subjective" sense another person *should* either fully believe the conclusion or fail to fully believe all the premises—even if we know that he doesn't recognize its validity (either because he's unaware of the proof or because he mistakenly rejects some of its principles). That doesn't rule out our employing other senses of 'should' (other kinds of constraints) that take account of his logical ignorance and that point in the other direction.

A somewhat similar issue arises from the fact that we may *think* an inference valid, but not be completely sure that it is. (Again, this could be either because we recognize our fallibility in determining whether complicated arguments are, say, classically valid, or because we aren't totally certain that *classically* valid arguments are really *valid*.) In that case, though we think the argument valid, there's a sense in which we should take account of the possibility that it isn't in deciding how firmly to believe a conclusion given that we fully believe the premises. But the solution is also similar: to the extent we think it valid, we think that there's a non-subjective sense in which we should either not fully believe the premises or else fully believe the conclusion; at the same time, we recognize that there are other senses of what we "should" believe that take more account of our logical uncertainty.

To summarize, we should qualify (VB) by saying that it concerns the core notion of validity, in which "extra" goals such as resource use are discounted; and also by adding that the notion of

[15] In an email exchange about MacFarlane (unpublished) that I discuss in Field 2009.

constraint (or ‘should’) in it is not univocal, and that (VB) is correct only on what I’ve called the non-subjective reading of the term.[16]

More interesting than the qualifications for (VB) is the need to expand it, and in three ways: (a) to cover not only full belief but also partial belief; (b) to cover not only belief (or acceptance) but also disbelief (or rejection); (c) to cover conditional belief. It turns out that these needed expansions interact in interesting ways. I’ll consider them in order.

2(a): Constraints on partial belief.

(VB) is stated in terms of full belief, but often we only have partial belief. How we generalize (VB) to that case depends on how we view partial belief. I’m going to suppose here that a useful (though at best approximate) model of this involves degrees of belief, taken to be real numbers in the interval [0,1]; so, representing an agent’s actual degrees of belief (credences) by Cr,

(1) $0 \leq \mathrm{Cr}(A) \leq 1$.

(I don’t assume that an agent has a degree of belief in every sentence of his language—that would impose insuperable computational requirements. We should understand (1) as applying when the agent has the credence in question.) I do *not* suppose that degrees of belief obey all the standard probabilistic laws, for any actual person’s system of beliefs is probabilistically incoherent. I don’t even suppose that a person’s degrees of belief *should* obey all the standard probabilistic laws. Obviously that would fail on senses of ‘should’ that take into account the agent’s computational limitations and faulty logical theory, but even for what I’ve called the non-subjective sense it is contentious: for instance, since it prohibits believing any theorem of classical sentential logic to degree less than 1, it is almost certainly objectionable if those theorems aren’t all really valid. (As we’ll soon see, it is also objectionable on some views in which all the classical theorems are valid, e.g. supervaluationism.)

What then do I suppose, besides (1)? The primary addition is

(VP) Our degrees of belief should (non-subjectively) be such that

[16] A still less subjective sense of ‘should’ is the sense in which we shouldn’t believe anything false. On this “hyper-objective” sense, (VB) is too weak to be of interest.

(2) If $A_1,...,A_n \Rightarrow B$ then $\mathrm{Cr}(B) \geq \Sigma_i \mathrm{Cr}(A_i) - n + 1$.

To make this less opaque, let's introduce the abbreviation Dis(A) for 1−Cr(A); we can read Dis as "degree of disbelief". Then an equivalent and more immediately compelling way of writing (2) is

(2_{equiv}) If $A_1,...,A_n \Rightarrow B$ then $\mathrm{Dis}(B) \leq \Sigma_i \mathrm{Dis}(A_i)$.

That (2_{equiv}) and hence (2) is a compelling principle has been widely recognized, at least in the context of classical logic: see for instance Adams 1975 or Edgington 1995.

But (2_{equiv}) and hence (2) also seem quite compelling in the context of non-classical logics, or at least many of them.[17] They explain many features of the constraints on degrees of belief typically associated with those logics.

To illustrate this, let's look first at logics that accept the classical principle of explosion

(EXP) $A \wedge \neg A \Rightarrow B$,

that contradictions entail everything. Or equivalently given the usual rules for conjunction,

(EXP*) $A, \neg A \Rightarrow B$.[18]

Since we can presumably find sentences B that it's rational to believe to degree 0, (2) applied to (EXP) tells us that Cr($A \wedge \neg A$) should always be 0, in the probability theory[19] for these logics as in classical probability theory; and (2) as applied to (EXP*) tells us that Cr(A) + Cr($\neg A$) shouldn't ever be greater than 1. These constraints on degrees of belief are just what we'd expect for a logic

[17] We would seem to need to generalize it somehow to deal with typical substructural logics. However, many logics that are often presented as substructural can be understood as obtained from an ordinary (non-substructural) logic by redefining validity in terms of a non-classical conditional in the underlying language. For instance, $A_1,...,A_n \Rightarrow_{\mathrm{substruc}} B$ might be understood as $\Rightarrow_{\mathrm{ord}} A_1 \rightarrow (A_2 \rightarrow ... (A_n \rightarrow B))$; or for other logics, as $\Rightarrow_{\mathrm{ord}} A_1 \circ (A_2 \circ ... (A_{n-1} \circ A_n)) \rightarrow B$, where $C \circ D$ abbreviates $\neg(C \rightarrow \neg D)$. I think these logics are best represented in terms of $\Rightarrow_{\mathrm{ord}}$, and that principle (2) in terms of $\Rightarrow_{\mathrm{ord}}$ is still compelling.

[18] (EXP*) is equivalent to disjunctive syllogism, given a fairly minimal though not wholly uncontentious background theory. (For instance, the argument from (EXP*) to disjunctive syllogism requires reasoning by cases.)

[19] Taking 'probability theory' to mean: theory of acceptable combinations of degrees of belief.

with these forms of explosion.

There are also logics that accept the first form of explosion but not the second. (This is possible because they don't contain ∧-Introduction.) The most common one is *subvaluationism* (the dual of the better-known supervaluationism, about which more shortly): see Hyde 1997. A subvaluationist might, for instance, allow one to simultaneously believe that a person is bald and that he is not bald, since on one standard he is and on another he isn't; while prohibiting belief that he is both since there is no standard on which he's both. On this view, it would make sense to allow the person's degrees of belief in A and in $\neg A$ to add to more than 1, while still requiring his degree of belief in $A \wedge \neg A$ to be 0: just what one gets from (2), in a logic with (EXP) but not (EXP*).

Let's also look at logics that accept excluded middle:

(LEM) $\Rightarrow A \vee \neg A$.

(2) tells us that in a logic with (LEM), $Cr(A \vee \neg A)$ should always be 1. Interestingly, we don't have a full duality between excluded middle and explosion, in the current context: there is no obvious (LEM)-like requirement that in conjunction with (2) leads to the requirement that $Cr(A) + Cr(\neg A)$ shouldn't ever be less than 1. For this we would need a principle (LEM*) that bears the same relation to (LEM) that (EXP*) bears to (EXP), but the notation of implication statements isn't general enough to formulate such a principle.

Indeed, one view that accepts excluded middle is supervaluationism,[20] and the natural way

[20] For supervaluationism, see Fine 1975. But note that while early in the article Fine builds into the supervaluationist view that what he calls supertruth is the same as truth, at the end of the article he suggests an alternative on which a new notion of determinate truth is introduced, and supertruth is equated with that, with "⟨A⟩ is true" taken to be equivalent to A. I think the alternative an improvement. But either way, I take the idea to be that supertruth rather than truth is the proper goal of belief, so that it may be allowable (and perhaps even compulsory) to fully believe a given disjunction while fully *dis*believing both disjuncts.

For instance, it would be natural for a supervaluationist who doesn't identify truth with supertruth to hold that neither the claim that a Liar sentence is true, nor the claim that it isn't true, is super-true; in which case it's permissible (and indeed in the case of non-contingent Liars, mandatory) to believe that such a sentence is either true or not true while disbelieving that it's true and disbelieving that it's not true. According to the supervaluationist, this wouldn't be ignorance in to model appropriate degrees of belief for supervaluationism is to allow degrees of belief in A and in $\neg A$ to add to less than 1 (e.g. when A is ‘Joe is bald’ and one considers Joe to be bald on some reasonable standards but not on others). More fully, one models degrees of belief in supervaluationist logic by Dempster-Shafer probability functions, which allow $Cr(A) + Cr(\neg A)$ to be anywhere in the interval [0,1] (while insisting that $Cr(A \vee \neg A)$ should be 1 and $Cr(A \wedge \neg A)$ should be 0).[21] Obviously this requires that we *not* accept the classical rule that we should have

(3?) $$Cr(A \vee B) = Cr(A) + Cr(B) - Cr(A \wedge B).$$

(We do require that the left hand side should be no less than the right hand side; indeed we require a more general principle, for disjunctions of arbitrary length, which you can find in the references in the first sentence of the previous note.) So it is unsurprising that we can’t get that $Cr(A) + Cr(\neg A) \geq 1$ from (LEM) all by itself.

If we do keep the principle (3?) in addition to (2), then if our logic includes (EXP) we must have

$$Cr(A \vee \neg A) = Cr(A) + Cr(\neg A),$$

and if our logic has (LEM) we must have

---

any normal sense: ignorance is a fault, and in this case what would be faulty (even in the non-subjective sense) would be to believe one of the disjuncts. (If I think I’m ignorant as to whether the Liar sentence is true, I’m not a supervaluationist; rather, I think that either a theory on which it isn’t true (e.g. a classical gap theory) is correct, or one on which it is true (e.g. a glut theory) is correct, but am undecided between them. The supervaluationist view is supposed to be an alternative to that.)

[21] See Shafer 1976; and for a discussion of this in the context of a supervaluationist view of vagueness (in the sense of supervaluationism explained in the previous footnote), Chapter 10 of Field 2001. (I no longer accept the supervaluationist view of vagueness or the associated Dempster-Shafer constraints on degrees of belief: see Field 2003.)

Some (e.g. Schiffer 2003; MacFarlane 2010) have thought that taking $Cr(A)$ and $Cr(\neg A)$ to be specific numbers that in the case of crucial vagueness and the paradoxes add to less than 1 doesn’t do justice to our feeling “pulled in both directions” by crucially vague and paradoxical sentences. I think this objection misfires: a view that represents our attitude by a pair of numbers $Cr(A)$ and $Cr(\neg A)$ that add to less than 1 can be equivalently represented by an assignment of the interval $[Cr(A), 1-Cr(\neg A)]$ to A and $[Cr(\neg A), 1-Cr(A)]$ to $\neg A$; and this latter representation does clear justice to our being “pulled both ways”.

$$\mathrm{Cr}(A \wedge \neg A) = \mathrm{Cr}(A) + \mathrm{Cr}(\neg A) - 1.$$

The first is part of the natural probability theory for "strong Kleene logic", the latter part of the natural probability theory for Priest's "logic of paradox" LP. Both these logics also take $\neg\neg A$ to be equivalent to A, which given (2) yields the additional constraint

$$\mathrm{Cr}(\neg\neg A) = \mathrm{Cr}(A).$$

Note that in the contexts of any of these logics (supervaluationist and subvaluationist as well as Kleene or Priest), we can keep the definition of "degree of disbelief" Dis as 1 minus the degree of belief Cr.[22] What we must do, though, is to reject the equation that we have in classical probability, between degree of disbelief in A and degree of belief in $\neg A$. In Kleene and supervaluational logic, Dis(A) (that is, $1-\mathrm{Cr}(A)$) can be greater than $\mathrm{Cr}(\neg A)$; in Priest and subvaluation logic it can be less.

The point of all this is to illustrate that (VP) (which incorporates Principle (2)) is a powerful principle applicable in contexts where we don't have full classical logic, and leads to natural constraints on degrees of belief appropriate to those logics. And (VP) is a generalization of (VB): (VB) is simply the very special case where all the $\mathrm{Cr}(A_i)$ are 1. (I've formulated (VP) in normative terms; the attitudinal variant is

$(\mathrm{VP})_a$ To regard the argument from $A_1,...,A_n$ to B as valid is to accept a constraint on degrees of belief: one that prohibits having degrees of belief where $\mathrm{Cr}(B)$ is less than $\Sigma_i\mathrm{Cr}(A_i) - n + 1$; i.e. where $\mathrm{Dis(B)} > \Sigma_i\mathrm{Dis}(A_i)$.)

2(b): Constraints on belief and disbelief together.

Let us now forget about partial belief for a while, and consider just full belief and full disbelief. Even for full belief and disbelief, (VB) is too limited. Indeed, it doesn't directly deal with disbelief at all. We can derive a very limited principle involving disbelief from it, by invoking the assumption that it is impossible (or at least improper) to believe and disbelieve

[22] If we do so, then the situation we have in Kleene logic and supervaluationist logic, that the degrees of belief in A and $\neg A$ can add to less than 1, can be equivalently put as the situation where the degrees of *dis*belief add to *more* than 1. Analogously, in Priest logic and subvaluation logic, the degrees of disbelief can add to *less* than 1.

something at the same time, but this doesn't take us far. Can we, without bringing in partial belief, build in disbelief in a more significant way? Yes we can, and doing so gives information that the probabilistic generalization above doesn't.

The idea here is (as far as I know) due to Greg Restall (2005). He proposes an interpretation of Gentzen's sequent calculus in terms of belief (or acceptance) and disbelief (or rejection). The idea is that the sequent $A_1,...,A_n \Rightarrow B_1,...,B_m$ directs you *not to fully believe all the* A_i *while fully disbelieving all the* B_j. (Restall doesn't explicitly say 'fully', but I take it that that's what he means: otherwise the classically valid sequent $A_1,...,A_n \Rightarrow A_1 \wedge ... \wedge A_n$ would be unacceptable in light of the paradox of the preface.) The idea doesn't depend on the underlying logic being classical. And it has the nice result that disbelief is built into the system completely on par with belief.

To illustrate, one of the principles in the classical sequent calculus is

(RC) $$A_1 \vee A_2 \Rightarrow A_1, A_2.$$

("Reasoning by cases".) On Restall's interpretation, this is a prohibition against fully believing a disjunction while fully disbelieving each disjunct; something which (VB) doesn't provide. Of course, one might reject that prohibition—supervaluationists do (on the interpretation offered in note 20, on which our goal should be to believe supertruths and to disbelieve things that aren't supertrue). But if one does so, one is rejecting the sequent (RC).

An apparently serious problem with Restall's proposal is that when applied to sequents with a single sentence in the consequent, it yields less information than (VB). That is, $A_1,...,A_n \Rightarrow B$ would direct us *not to fully reject* B while fully accepting $A_1,...,A_n$; whereas what (VB) directs, and what I assume we want, is *to fully accept* B whenever one fully accepts $A_1,...,A_n$, at least if the question of B arises. In other words, if $A_1,...,A_n \Rightarrow B$ (and one knows this), then to fully accept all of $A_1,...,A_n$ while *refusing to accept* B seems irrational; but the Restall account fails to deliver this. By contrast, the approach in terms of partial belief, offered in Section 2(a), handles this: it tells us that when the $Cr(A_i)$ are 1, $Cr(B)$ should be too.

Should we then abandon Restall's approach for the partial belief approach? When I first read Restall's paper, that's what I thought. But I now see that the proper response is instead to combine them—or rather, to find a common generalization of them.

2(c). The synthesis.

What we need is to generalize the formula (2) to multiple-conclusion sequents. The best way to formulate the generalized constraint, so as to display the duality between belief and disbelief, is to either use degrees of belief Cr for sentences in the antecedent of a sequent and degrees of disbelief (Dis = 1−Cr) for sentences in the consequent, or the other way around. In this formulation, the constraint is: our degrees of belief should satisfy

(2^+) If $A_1,...,A_n \Rightarrow B_1,...,B_m$ then $\Sigma_{i\leq n}\,\mathrm{Cr}(A_i) + \Sigma_{j\leq m}\,\mathrm{Dis}(B_j) \leq n + m - 1$;

or equivalently,

If $A_1,...,A_n \Rightarrow B_1,...,B_m$ then $\Sigma_{i\leq n}\,\mathrm{Dis}(A_i) + \Sigma_{j\leq m}\,\mathrm{Cr}(B_j) \geq 1$.

(That's the normative form, call it $(\mathrm{VP}^+)_n$. The attitudinal form is:

$(\mathrm{VP}^+)_a$ To regard the sequent $A_1,...,A_n \Rightarrow B_1,...,B_m$ as valid is to accept (2^+) as a constraint on degrees of belief.)

Note the following:

(i) When there's only one consequent formula, the right hand side of (2^+) reduces to

$$\mathrm{Cr}(A_1) + ... + \mathrm{Cr}(A_n) + 1 - \mathrm{Cr}(B) \leq n,$$

so (2^+) yields precisely the old (2).

(ii) When we fully reject each B_j, i.e. when each $\mathrm{Dis}(B_j)$ is 1, the right hand side of (2^+) yields

$$\mathrm{Cr}(A_1) + ... + \mathrm{Cr}(A_n) \leq n - 1;$$

that is,

$$\mathrm{Dis}(A_1) + ... + \mathrm{Dis}(A_n) \geq 1.$$

(iii) As a special case of (ii), when we fully reject each B_j and fully accept $n-1$ of the antecedent formulas, you must *fully* reject the other one. This is a stronger version of Restall's constraint, which was that you shouldn't fully reject each B_j while fully accepting each A_i. So (2^+) generalizes Restall's constraint as well as generalizing (2).

These seem to be intuitively the right results, and I leave it to the reader to convince him- or herself that the results in every other case are just what we should expect.

Because this formulation yields that in Section 2(a) as a special case, it is immediate that it resolves the problem of excessive weakness that I raised in Section 2(b) for using the unprobabilized sequent calculus as the desired constraint on belief and disbelief.

And because it also yields Restall's constraint as a special case, it also has the advantage that that account has over the simpler probabilistic account: it yields the constraint on our attitudes that the acceptance of reasoning by cases provides.[23]

And indeed it does considerably better than the unprobabilized Restall on this. Recall that in a sequent formulation, reasoning by cases is represented by the sequent

(RC) $A_1 \vee A_2 \Rightarrow A_1, A_2$.

On Restall's constraint, this means that we shouldn't fully accept a disjunction while fully rejecting each disjunct—information that the old constraint (2) of Section 2(a) didn't provide. But the generalized constraint tells us still more: it tells us that

$$\mathrm{Cr}(A_1 \vee A_2) \text{ should be less than or equal to } \mathrm{Cr}(A_1) + \mathrm{Cr}(A_2).$$

Recall that that's a constraint accepted in the probability theory for Kleene logic (it follows from (3?) plus (2) plus (EXP)). But the constraint is rejected in the probability theory for supervaluationist logic (i.e. the Dempster-Shafer theory): the latter even allows $\mathrm{Cr}(A_1 \vee A_2)$ to be 1 when $\mathrm{Cr}(A_1)$ and $\mathrm{Cr}(A_2)$ are each 0. (Thus if A_1 is the claim that a Liar sentence is true and A_2 is the negation of A_1, a supervaluationist demands that the disjunction be fully accepted, but will allow that both disjuncts be fully rejected—indeed, will probably demand this, in the case of a non-

[23] At least, this is so for reasoning by cases in the form that takes you from $\Gamma,A\Rightarrow X$ and $\Delta,B\Rightarrow Y$ to $\Gamma,\Delta,A\vee B\Rightarrow X,Y$. To get the form that takes you from $\Gamma,A\Rightarrow X$ and $\Gamma,B\Rightarrow X$ to $\Gamma,A\vee B\Rightarrow X$ one also needs structural contraction. But the generalization to be considered in Section 2(d) delivers that.

contingent Liar.) The reason this can happen in supervaluationist logic is that that logic rejects (RC). The generalized constraint thus shows how this dramatic difference between two probability theories that allow $Cr(A) + Cr(\neg A)$ to be less than 1 arises out of a difference in their underlying logic.

2(d). Conditional belief.

There is an obvious strengthening of (VP) and (VP^+) that I've been suppressing: we should make them principles governing not only belief, but also conditional belief. For instance, we might strengthen (VP^+) by strengthening (2^+) to

$$\text{If } A_1,...,A_n \Rightarrow B_1,...,B_m \text{ then for all } C, \Sigma_{i\leq n} \text{Dis}(A_i|C) + \Sigma_{j\leq m} \text{Cr}(B_j|C) \geq 1.$$

Part of my reason for not having made this strengthening earlier is to show how much can be done without it. Another part of the reason is that conditional degree of belief is more complicated in nonclassical logics than in classical. Indeed, since in typical nonclassical logics a distinction is made between rejection and acceptance of the negation, the simple form of conditional credence $Cr(A|C)$ is inadequate: $Cr(A|C)$ is the credence in A conditional on full acceptance of C, but we'll need a more general notion $Cr(A|C/D)$ of the credence in A conditional on full acceptance of C and full rejection of D.[24] (In classical logic this will just be $Cr(A|C \wedge \neg D)$, but if rejection isn't just acceptance of the negation they will differ.) So the above strengthening of (2^+) should be adjusted:

(2^+_{cond}) $$\text{If } A_1,...,A_n \Rightarrow B_1,...,B_m \text{ then for all } C \text{ and } D, \Sigma_{i\leq n} \text{Dis}(A_i|C/D) + \Sigma_{j\leq m} \text{Cr}(B_j|C/D) \geq 1.$$

[24] The usual central law of conditional probability, generalized to the extra place, is

$$Cr(A \mid B\wedge C \,/\, D)\cdot Cr(B \mid C/D) = Cr(A\wedge B \mid C/D).$$

We impose also the dual law

$$Dis(A \mid C/\, B\vee D)\cdot Dis(B \mid C/D) = Dis(A\vee B \mid C/D).$$

Some policy is needed for how to understand $Cr(A \mid C/D)$ and $Dis(A \mid C/D)$ when the logic dictates that C can't be coherently accepted or D can't be coherently rejected; I tentatively think the best course is to let both $Cr(A \mid C/D)$ and $Dis(A \mid C/D)$ be 1 in these cases. Spelling it out, the idea would be to call A *absurd* if for all X and Y, $Cr(A|X/Y) = 0$, and *empty* if for all X and Y, $Dis(A|X/Y) = 0$; then the requirement that $Dis(A \mid C/D) = 1 - Cr(A \mid C/D)$ holds only for *acceptable pairs* $<C,D>$, pairs where C isn't absurd and D isn't empty. (Some derived laws will then be subject to the same restriction.) But this is only one of several possible policies for dealing with these "don't care" cases.

While as I've illustrated, much can be done to illuminate the connection between validity and degrees of belief using only (2^+), one needs (2^+_{cond}) to go much further. For instance, one thing I haven't done is to use generalized probabilistic constraints to derive (3?) from the logical principles of those logics where (3?) is appropriate (e.g. Kleene logic and LP, as well as classical). With (2^+_{cond}) we can do this. This follows from two more general consequences of (2^+_{cond}): first, that for logics with ∧-Introduction (and a few other uncontroversial laws),

(I) $\quad Cr(A|C/D) + Cr(B|C/D) \leq Cr(A \wedge B|C/D) + Cr(A \vee B|C/D)$;

second, that for logics with ∨-Elimination (and a few other uncontroversial laws),

(II) $\quad Cr(A|C/D) + Cr(B|C/D) \geq Cr(A \wedge B|C/D) + Cr(A \vee B|C/D)$.

I leave the proofs to a footnote.[25] I'd conjecture that (2^+_{cond}) is enough to get from any logic (of the kinds under consideration here) to *all* reasonable laws about credences appropriate to that logic.

I've been talking about how to derive constraints on degrees of belief from the logic, but it would be natural to argue that conversely, we can obtain the logic from the probability theory by turning a principle in the general ballpark of (2^+) into a biconditional. But the converse of (2^+_{cond}) is far more plausible than that of (2^+); moreover, we need it to be (2^+_{cond}) rather than (2^+) to carry out the idea technically. For to fully derive the logic, one of the things one needs to derive is the Gentzen structural rules. Of these, Reflexivity, Thinning and Permutation are immediately obvious

---

[25] ∧-I gives: $\quad Cr(A \mid C \wedge (A \vee B) / D) + Cr(B \mid C \wedge (A \vee B) / D) \leq Cr(A \wedge B \mid C \wedge (A \vee B) / D) + 1$.

Multiplying each term by $Cr(A \vee B \mid C / D)$, and using the central law of conditional probability (note 24), we get

$$Cr(A \wedge (A \vee B) \mid C / D) + Cr(B \wedge (A \vee B) \mid C / D) \leq Cr((A \wedge B) \wedge (A \vee B) \mid C / D) + Cr(A \vee B \mid C / D).$$

Using obvious equivalences this yields (I).

∨-I gives: $\quad Dis(A \vee B \mid C / D \wedge (A \wedge B)) + 1 \geq Dis(A \mid C / D \wedge (A \wedge B)) + Dis(B \mid C / D \wedge (A \wedge B))$.

Multiplying each term by $Dis(A \wedge B \mid C / D)$, and using the dualized central law of conditional probability (note 24), we get

$$Dis((A \vee B) \vee (A \wedge B) \mid C / D) + Dis(A \wedge B \mid C / D) \geq Dis(A \vee (A \wedge B) \mid C / D) + Dis(B \vee (A \wedge B) \mid C / D).$$

Using obvious equivalences this yields:

$$Dis(A \mid C / D) + Dis(B \mid C / D) \leq Dis(A \vee B \mid C / D) + Dis(A \wedge B \mid C / D),$$

which yields (II).

from the biconditional form of either (2^+) or (2^+_{cond}). The same holds for Cut in the form

$$\frac{\Sigma_1 \Rightarrow \Delta_1, C \qquad \Sigma_2, C \Rightarrow \Delta_2}{\Sigma_1,\Sigma_2 \Rightarrow \Delta_1,\Delta_2.}$$

But structural contraction requires the generalization over C and D in (2^+_{cond}): one must vary the C in the biconditional form of (2^+_{cond}) to get contraction on the left, and the D to get contraction on the right.[26]

Once one has the structural rules, there is no problem getting the other logical meta-rules appropriate to given laws of credence: as remarked above, reasoning by cases then falls out whenever the laws on credences include

$$\mathrm{Cr}(A_1 \vee A_2 \mid C/D) \leq \mathrm{Cr}(A_1 \mid C/D) + \mathrm{Cr}(A_2 \mid C/D)$$

(see note 23), and it's easy to see that conditional proof falls out whenever they include that, $\mathrm{Cr}(A \vee \neg A \mid C/D) = 1$, and $\mathrm{Cr}(A \rightarrow B \mid C/D) \geq \max\{\mathrm{Cr}(\neg A \mid C/D), \mathrm{Cr}(B \mid C/D)\}$.

---

[26] We also need that $\mathrm{Cr}(A \wedge A \mid C/D) = \mathrm{Cr}(A \mid C/D) = \mathrm{Cr}(A \vee A \mid C/D)$ and $\mathrm{Cr}(A \wedge B \mid C/D) \leq \mathrm{Cr}(A \mid C/D) \leq \mathrm{Cr}(A \vee B \mid C/D)$; the first of these and the laws in note 24 yield (i) $\mathrm{Cr}(A \mid C \wedge A / D) = 1$ and (ii) $\mathrm{Cr}(A \mid C / D \vee A) = 0$.

Contraction on left (leaving out side formulas for simplicity): To establish the meta-rule $A, A \Rightarrow B / A \Rightarrow B$, we must get from

$$\forall C \forall D[\mathrm{Cr}(B \mid C/D) \geq 2\mathrm{Cr}(A \mid C/D) - 1]$$

to

$$\forall C \forall D[\mathrm{Cr}(B \mid C/D) \geq \mathrm{Cr}(A \mid C/D)].$$

But there's no problem doing so: rewrite the bound C in the assumption as C^*, and instantiate it by $C \wedge A$, to get $\forall C \forall D[\mathrm{Cr}(B \mid C \wedge A / D) \geq 2\mathrm{Cr}(A \mid C \wedge A / D) - 1]$, which by (i) yields $\forall C \forall D[\mathrm{Cr}(B \mid C \wedge A / D) = 1]$. Multiplying both sides by a common factor we get $\forall C \forall D[\mathrm{Cr}(B \mid C \wedge A / D) \cdot \mathrm{Cr}(A \mid C/D) = \mathrm{Cr}(A \mid C/D)]$, which by laws in note 24 again yields $\forall C \forall D[\mathrm{Cr}(A \wedge B \mid C/D) = \mathrm{Cr}(A \mid C/D)]$, and the left hand side is $\leq \mathrm{Cr}(B \mid C/D)$; so $\forall C \forall D[\mathrm{Cr}(A \mid C/D) \leq \mathrm{Cr}(B \mid C/D)]$, as desired.

Contraction on the right (again leaving out side formulas): To establish the meta-rule $A \Rightarrow B,B / A \Rightarrow B$, we must get from

$$\forall C \forall D[2\mathrm{Cr}(B \mid C/D) \geq \mathrm{Cr}(A \mid C/D)]$$

to

$$\forall C \forall D[\mathrm{Cr}(B \mid C/D) \geq \mathrm{Cr}(A \mid C/D)].$$

Rewrite the bound D in the assumption as D^*, and instantiate it by $D \vee B$, to get $\forall C \forall D[2\mathrm{Cr}(B \mid C / D \vee B) \geq \mathrm{Cr}(A \mid C / D \vee B)]$, which by (ii) yields $\forall C \forall D[\mathrm{Cr}(A \mid C / D \vee B) = 0]$, i.e. $\forall C \forall D[\mathrm{Dis}(A \mid C / D \vee B) = 1]$. Multiplying both sides by a common factor we get $\forall C \forall D[\mathrm{Dis}(A \mid C / D \vee B) \cdot \mathrm{Dis}(B \mid C/D) = \mathrm{Dis}(B \mid C/D)]$, which by laws of note 24 yield $\forall C \forall D[\mathrm{Dis}(A \vee B \mid C/D) = \mathrm{Dis}(B \mid C/D)]$, and the left hand side is $\geq \mathrm{Cr}(A \mid C/D)$; so $\forall C \forall D[\mathrm{Cr}(A \mid C/D) \leq \mathrm{Cr}(B \mid C/D)]$, as desired.

Details aside, I think this discussion shows that it is illuminating to view validity as providing constraints on our (conditional) degrees of belief. And debates about validity are in effect debates about which constraints on (conditional) degrees of belief to adopt. The next two sections deal with some issues about how to understand this.

3. Normative definitions or primitivism?

The view I've been advocating has it that instead of trying to define validity in other terms, as the necessary truth-preservation account does, we should take it as a primitive, and explain its conceptual role in terms of how it constrains our (conditional) beliefs.

We should contrast this with another proposal: that we define validity, but in normative terms, in a way that reflects the connection between validity and belief. That alternative proposal is that we define $A_1,...,A_n \Rightarrow B$ as "One shouldn't (in the non-subjective sense) fully believe $A_1,...,A_n$ without fully believing B"; or a variant of this based on (VP) or (VP^+) or (VP_{cond}) or (VP^+_{cond}) instead of (VB). I won't attempt a serious discussion of this, but I think it would sully the purity of logic to define validity in normative terms whose exact content is less than clear.

But is the approach in Section 2 significantly different? I think so. Compare the notion of chance. I take it that no one would want to say that a claim such as

The chance of this atom decaying in the next minute is 0.02

is equivalent in meaning to

You ought to believe to degree 0.02 that the atom will decay in the next minute.

The first claim, unlike the second, is not literally about what we should believe. On the other hand, there seems to be no prospect of a reductive account of chance. And it seems clear that an important part of our understanding of the first claim lies in its conceptual ties to the second: what would we make of someone who claimed to accept the first claim but thought that it was rational to strongly disbelieve that the atom will decay in the next minute? While claims about chance aren't literally claims about what we should believe, it's hard to deny that debates about the chance of an atom decaying in the next minute are intimately related to debates about what degree of belief to have in such a decay. I think the situation with validity is much like that with chance.

In the chance case as in the validity case, the "ought" in question is in an important sense "non-subjective": it does not take into consideration the agent's evidence. If the chance of decay is actually 0.02, but a person has reason to think it is far higher, then there's a subjective sense of 'ought' in which the person ought to believe in decay to that higher degree, not to degree 0.02. But as in the case of validity, the existence of the idea of "ought based on one's evidence" doesn't drive out the existence of a non-subjective ought, in which one ought to believe in decay to degree 0.02.

It seems, then, that a large part of our understanding of the notion of chance is our acceptance of something like the following principle:

$(C)_n$ If the chance of *A* is *p*, then (in the non-subjective sense of 'ought') our degree of belief in *A* ought to be *p*.

A more accurate formulation would be:

$(C_{cond})_n$ If the chance of *A* happening in situation *B* is *p*, then (in the non-subjective sense of 'ought') our conditional degree of belief in *A* given *B* ought to be *p*.

That something like this is so seems relatively uncontroversial.

$(C)_n$ and its variant $(C_{cond})_n$ seem closely analogous in form to the $(V)_n$ principles that I have suggested for validity. It seems to me that for validity as for chance, a primitivist (i.e. non-definitional) conceptual role account, rather than either a non-normative reduction or a normative reduction, is the way to go, and that it is no more problematic for validity than it is for chance.

4. Realism vs. projectivism.

There is, I admit, a worry about the primitivist line: not just in the case of validity, but in the case of chance as well. The worry is that the $(C)_n$ principles and the $(V)_n$ principles employ non-subjective 'ought's, and one might reasonably wonder what sense is to be made of them. If someone were to ask what's wrong with having degrees of belief that violate one of these principles, is there no answer to be made beyond saying in one's most serious voice "*It is forbidden!*"?

Suppose I know that the chance of an atom decaying in the next minute, given its nature and circumstances, is 0.02. Naturally enough, I believe to degree 0.02 that it will so decay; but Jones believes to degree 0.9 that the atom will decay in the next minute. (He might have the corresponding view about the chances, but needn't: e.g. he might think it isn't a chance process.[27])

[27] Or more fancifully, there might be an odd disconnect between his belief about the chance and his degree of belief about the decay.

Even if his degree of belief in the decay is reasonable given his evidence, there seems to be a sense in which it's wrong given the actual chances—and that's so even if the atom *does* decay in the next minute.

But if asked *what's* wrong with having a degree of belief that doesn't accord with the chances, what can I say? To say "There's a high chance of what one believes being false" is obviously unhelpful: why is the *chance* of its being false, as opposed to its actually being false, relevant? It is also no help to say

> "What's wrong is that if he had a policy of betting in terms of that degree of belief in similar situations, *the chance is very high that* he'd lose money".

There are two reasons for this, both familiar. The first is relevance: the person with degree of belief 0.9 in this instance needn't have a long term policy of believing to degree 0.9 in all similar instances, so why would a problem with the long term strategy imply a problem with the individual instance? But the more basic reason is that the imagined answer merely pushes the question back: what's wrong with Jones thinking he'll win with this long-term betting strategy, given merely that the actual *chance* of winning is low? It seems obvious that there is something wrong with it, even if it was reasonable given his evidence and in addition he happens to be very lucky and wins; but that's just another instance of the idea that our degrees of belief ought to correspond to the chances, which is what it was attempting to explain.

So this attempt to explain the sense in which there's something wrong with degrees of belief that don't accord with the chances is hopeless. If there's no possibility of doing better, we will have to either give up some natural 'ought' judgements or settle for a primitive non-subjective ought on which nothing more can be said than that you are forbidden to believe to any degree other than 0.02. But *who* forbids it, and why should we care?

There have been various closely related proposals for a more informative answer to the "What's wrong?" question in the case of chance. (See for instance Jeffrey 1965, Blackburn 1980 and Skyrms 1984.) Roughly put, the common theme is that we regard chances as projections from our epistemic state. They aren't just credences of course, but rather "de-subjectivized credences".[28] The de-subjectivization means that the 'ought's based on them don't merely involve

[28] 'Credence' as often understood is already a somewhat de-subjectified notion, in that constraints

what an agent's degrees of belief ought to be *given her evidence, beliefs, etc.*, but are quite a bit more objective than that; still, the chances are ultimately based on credences, in a way that makes the corresponding 'ought's unmysterious.

I think that the general idea has considerable appeal, and what I want to suggest is that an analogous idea has a similar appeal in the case of validity. Indeed, it seems to me that whatever difficulties the idea has for the case of validity it equally has for the case of chance. I will not attempt a precise formulation, in either the case of chance or validity: indeed I'm deliberately presenting the general idea in such a way as to allow it to be filled out in more than one way. The goal of the section is simply to pursue the analogy between the two cases. (And I should say that while I'm sympathetic to the projectivist view in both cases, nothing outside this section of the paper requires it.)

In the case of chance, the projectivist idea involves temporarily shifting the subject from chance itself to attributions of chance. Instead of directly asking

(i) What is it for there to be chance p that under conditions B, A will happen?

we start out by asking

(ii) What is it for someone *to take* there to be chance p that under conditions B, A will happen?

We then argue that to answer (ii), *we needn't use the notion of chance to characterize the content of the person's mental state*. (This is the first stage of the "projectivism".) For instance, the answer to (ii) might be that the person has a kind of resilient conditional degree of belief in A given B (and perhaps, that he recommends this conditional degree of belief to others): for any C of a certain kind, $Cr(A \mid B \wedge C) = p$.[29] The restriction that C be "of a certain kind" is hard to spell out precisely, but a very rough stab might be that C contains no information about times later than the conditions referred to in B.[30] This is only very rough: one important reason for this is that evidence

---

of coherence are imposed, but the idea is to de-subjectivize it further in a way that makes it less dependent on evidence.

[29] Alternatively, it might be that the conditional degree of belief that best fits his overall system of degrees of belief is like this, even though his actual degree of belief is different.

[30] The issue is similar to the issue of clarifying "admissible evidence" in Lewis's "Principal Principle": see Lewis 1980.

about past experiments might lead one to alter our physical theory in a way that led us to change our conditional probabilities, and this should count as *changing our views about what the chances are* rather than *showing that our conditional probabilities weren't resilient enough to count as views about chance*. I doubt that the idea of resilience can be made perfectly precise, but also doubt that we need total precision for the notion to be useful in an account of chance-attribution.

In short, the idea is that the concept of chance is illuminated by the following principle:

$(C_{cond})_a$ To regard the chance of *A* happening, under condition *B*, as *p* is to accept a constraint on belief: it is to demand the conditional degree of belief *p* in *A*, given *B* and *C*, for any admissible *C*.

This of course is broadly analogous to our $(V)_a$ principles for validity: those characterize what it is for someone to take an argument to be valid without using the notion of validity to characterize the content of her mental state.

$(C_{cond})_a$ explains what would be wrong with someone *taking* the chance of *A* in situation *B* to be 0.02 and yet (despite having no other relevant information) knowingly having credence 0.9 of *A* given *B*: such a person would be seriously unclear on the concept of chance. Similarly, the $(V)_a$ principles explain what would be wrong with someone (e.g. Lewis Carroll's Achilles) taking modus ponens to be valid in all instances, and yet (in full awareness of what he was doing) fully accepting the premises of a modus ponens while refusing to accept its conclusion.

Does this account of chance-*judgements* and validity-*judgements* shed any light on what chance and validity *are*? The key second stage of projectivism is to argue that the answer is yes. The idea isn't to reject the normative principles $(C_{cond})_n$ and $(V)_n$; rather, these principles properly express the projectivist's attitudes. But the projectivist will see the acceptance of $(C_{cond})_n$ as founded on the more basic truth $(C_{cond})_a$, and take this to remove worries about the 'ought's appealed to in $(C_{cond})_n$; and similarly for $(V)_n$.

I won't try to spell out this second stage of the projectivist account; there are different ways in which the details could go. But in the chance case, it should be clear that if I myself have a resilient conditional degree of belief *p* in *A* given *B*, and know that someone else has evidence that supports a very different conditional degree of belief *q*, I can distinguish between what her conditional degree of belief should be given her evidence (viz., *q*) and what it "objectively" should be (viz., *p*, ignoring the possibility that I myself am misled). Similarly in the validity case: if I myself believe in the law of explosion but I know that a friend doesn't, I can distinguish between what she should believe about a matter *given her logical views* and what she "objectively" should believe about it. Judging what she "objectively" should believe about validity is straightforward on the view (at least if I ignore the possibility that I myself am misled): it's the same as judging what's valid. Judging what she should believe *given her logical views* involves some sort of projection, something like an "off-line decision" of what to believe about the matter in question on the pretense that one's views are in key respects like hers.

The distinction between what X should believe given his or her current evidence or beliefs and what he or she "objectively" should believe is equally unproblematic when X is my past self, or my counterpart in another possible world. And it isn't a whole lot more problematic even when X is my actual current self. If I have a resilient conditional degree of belief *p* in *A* given *B*, and am sure that it's appropriate given my evidence, I can still make sense of the possibility that despite my evidence, the chance is different: in doing so I am projecting to a different epistemic state than the one I now occupy, perhaps one that might be obtained were I to learn new truths about frequencies or underlying mechanisms, or unlearn some falsehoods about these things. Similarly, such a view for validity can allow for a divergence between what my degrees of belief ought to be *given my current opinions about validity* and what they *"objectively"* ought to be: I can project to a dimly-imagined epistemic state where I learn things about logical space that would lead me to revise some of my current epistemic policies.

Similarly, I can make sense of both unknown chances and unknown validities: I'm projecting to judgements I'd make where I acquire additional knowledge of certain sorts.

In the last few paragraphs I've been putting 'objectively' in quotes, because there's a question as to whether this rather vague term is fully appropriate. And my view is that the "level of objectivity" we get is greater in some cases than in others. If I differ with someone on the chance of rain tomorrow, it's very likely that this difference is *largely* due to differences in information we have (information expressible without the term 'chance'); to the extent that this is so, there is certainly no threat to objectivity. There is also no threat to what I'd call objectivity if his belief differs from mine not because of a difference in evidence but because he takes seriously the predictions in some ancient text, and his taking such predictions seriously is impervious to evidence. In such a case, there are strong advantages for my conditional degrees of belief over his, and that seems to me to give all the "objectivity" we should ask for. (Of course he is likely to disagree with me about the advantages. In some cases that further disagreement might be based on differences in background information, but even if it isn't, I don't think any interesting notion of objectivity will count this case non-objective.)

Objectivity is a flexible notion. In the case of chance, it does not seem unreasonable to stipulate that any attribution of chance that departs drastically from actual stable frequencies is "objectively wrong", whatever the attributor's degrees of belief: one can say that the actual stable frequencies give a "metaphysical basis" for ruling out projections based on some credences. Still, the impossibility of an actual reduction of chances to frequencies means that there is some limitation on the "objectivity" achievable in this way. It is hard to find much non-objectivity about chance in examples such as radioactive decay, where the frequencies are so exceptionally stable; but most of us also speak of chance in other cases, such as coin tosses or weather patterns. And here, it is far from obvious that all disagreements about the chance of rain are fully explainable on the basis of either difference of evidence or gross irrationality (as it was in the ancient texts example). If not, I don't think it unreasonable to declare some such disagreements "non-objective". The "projectivism" I'm suggesting for chance has it that a judgement of chance carries with it a judgement of what a person ought to believe under appropriate circumstances, in a non-subjective sense, i.e. independent of what evidence the person has; but it may be that *to whatever extent person-independent facts such as frequencies don't settle the chances*, there is nothing to legitimize the judgement over various alternatives.

The role that frequencies play for chance can be played by truth-preservation for validity. Just as an attribution of chance that departs drastically from actual frequencies seems objectively wrong, an attribution of validity to an argument that definitely fails to preserve truth seems objectively wrong. (Recall (C) at the end of Section 1.) *Prima facie*, this suffices for making most disputes about validity objective; though there are issues about sameness of meaning across logical theories that might undermine that claim. But even putting those issues aside, the impossibility of an actual reduction of validity to logically necessary truth preservation is a sign that there may be *some* degree of non-objectivity in the choice of logic.

It may be, for instance, that a view that locates the failure of the Curry argument in modus ponens and a view that locates it in conditional proof can't be distinguished in terms of how closely validity corresponds to truth preservation. I don't say that this is the actual situation, but suppose it is. In that case, the difference between the views is irreducibly a matter of normative policy. The proponent of unrestricted modus ponens will say that we ought to conform our degrees of belief to it, in the sense I've described, and the proponent of unrestricted conditional proof will say that we ought to conform our degrees of belief to a different standard. And each will take their 'oughts' to be non-subjective, in the sense that they aren't merely claims about what we ought to do given our logical theory. That wouldn't be a serious lack of objectivity if there were strong advantages to one view over the other; but it may not be entirely obvious that there are. Perhaps each party in the dispute recognizes that despite his personal preference for one route over the other, there is no really compelling advantage on either side. (Again, I don't say that this is the actual situation, but suppose it is.) Given Curry's paradox, we can't declare *both* valid, without restricting central principles of truth; and let us assume for the sake of argument that restricting the truth principles in the required way would have far worse disadvantages than restricting either modus ponens or conditional proof. Moreover, since a logic with neither principle would be far too weak for serious use, only Buridan's ass would use the symmetry of the situation to argue that *neither* is valid. And so, in the circumstances I'm imagining, where there really is no strong advantage for one or the other, each party to the dispute uses whichever of the principles he or she is more comfortable with, while recognizing that the choice isn't objective in any serious sense.

In cases like this where we recognize that we need to be permissive about alternatives, our validity judgements will presumably be somewhat nuanced. If my logic of choice includes modus ponens, I will typically evaluate arguments in accordance with it; but in evaluating the arguments of someone who has a worked out system which restricts modus ponens to get conditional proof, it will sometimes be appropriate to project to that person's point of view. Sometimes, not always: e.g. if I find another person's arguments for an implausible conclusion prima facie compelling, I'll want to figure out where they go wrong on my own standards. In some situations it's best to explicitly relativize to one standard over the other (though at least if the non-objectivity is sufficiently widespread it is not possible to always do this: there wouldn't be enough of a common logic to be able to draw conclusions about the relativized concepts). I think that projectivism gets all of this right. But this is not the place to explore it further.[31]

5. Soundness and completeness.

I argued at the very start of this paper that real validity is neither model-theoretic nor proof-theoretic, and have been offering an account of what it is instead. This has an impact on how to understand soundness and completeness proofs.

What are usually called soundness and completeness results for logics are results that relate the proof theory to the model theory. Suppose we have a system S of derivations for, say, classical sentential logic, which enables us to define a derivability relation $\Gamma \vdash_S B$ (or a multiple conclusion analog $\Gamma \vdash_S \Delta$). Suppose we also have a model theory M for that logic, say the usual one in terms of 2-valued truth tables, which enables us to define a model-theoretic validity relation $\Gamma \vDash_M B$ (or $\Gamma \vDash_M \Delta$). Then the usual soundness and completeness theorems relate these. In what I consider the best case scenario,[32] these go as follows (focusing on the single consequent case, but the generalization to multiple consequent is obvious):

---

[31] The "projectivism" I'm advocating for deductive validity seems more obviously attractive in the case of inductive goodness. (1) Whereas those who ignore the paradoxes (and vagueness) might think that deductive validity is just logically necessary preservation of truth, no one could think that in the inductive case. For many reasons, the thought that one could find a notion of reliability that does an analogous job doesn't withstand scrutiny. (The main reason, I think, is that there simply is no useful notion of reliability for methods that "self-correct" in the way that induction does: see Section 4 of Chapter 13 of Field 2001.) (2) Whereas in the deductive case the idea that there is some degree of non-objectivity about which logic to adopt may seem surprising, it is far less so in the inductive case: any attempt to formulate an inductive method in detail will show that there are a large number of parameters affecting such matters as how quickly one adjusts ones degrees of beliefs on the basis of observed frequencies; the idea that there's exactly one "right" value to such a parameter seems absurd.

[32] For some logics it is common to employ proof procedures for which formal soundness must be either stated in a more complicated way than below, or else restricted to the case where Γ is empty. (For instance, proof procedures containing a generalization rule for the universal quantifier or a necessitation rule for a modal operator.) Use of such a proof procedure often goes with defining entailment in terms of logical truth, something which isn't possible in every logic. It seems to always be possible, though, to convert a proof procedure that is formally sound only in the more complicated or restricted sense to one of intuitively equivalent power which is formally sound in the sense given below.

Formal Soundness Theorem: For any set Γ of sentences and any sentence *B*, if $\Gamma \vdash_S B$ then $\Gamma \vDash_M B$.

Formal Completeness Theorem: For any set Γ of sentences and any sentence B, if $\Gamma \vDash_M B$ then $\Gamma \vdash_S B$.

But calling these theorems "soundness theorems" and "completeness theorems" seems slightly misleading, for they merely connect two different notions about each of which it could be asked whether they are sound or complete. What we really want to know is whether the model theory is sound and complete with respect to the validity relation, and whether the proof theory is sound and complete with respect to the validity relation; and neither of these questions is directly answered by the formal soundness and completeness theorems. Thus (using $\Gamma \Rightarrow B$ to mean that the argument from Γ to *B* is logically valid, as understood in the "primitivist" way suggested earlier in the paper)[33] we can formulate genuine soundness and completeness as follows:

(P-Sound) [Genuine Soundness of the proof theory]: For any set Γ of sentences and any sentence *B*, if $\Gamma \vdash_S B$ then $\Gamma \Rightarrow B$.

(P-Comp) [Genuine Completeness of the proof theory]: For any set Γ of sentences and any sentence *B*, if $\Gamma \Rightarrow B$ then $\Gamma \vdash_S B$.

(M-Sound) [Genuine Soundness of the model theory]: For any set Γ of sentences and any sentence *B*, if $\Gamma \vDash_M B$ then $\Gamma \Rightarrow B$.

(M-Comp) [Genuine Completeness of the model theory]: For any set Γ of sentences and any sentence *B*, if $\Gamma \Rightarrow B$ then $\Gamma \vDash_M B$.

These four soundness and completeness claims involve a notion ⇒ of validity that we're taking to be undefined, so there's no question of formally proving these "genuine" soundness and completeness claims. But to what extent can we convincingly argue for them nonetheless? The question, and the answer to follow, is I think a generalization of a question asked and answered in Kreisel 1967. (Kreisel's discussion is sometimes understood as solely concerning "set-sized interpretations" of quantification theory versus "interpretations with domains too big to be a set";

[33] 'Logically valid' is actually a contextually relative notion. If it's sentential logic that is in question, then we should take 'logically valid' here to mean 'valid on the basis of sentential structure'; if quantificational logic, 'valid on the basis of quantificational structure'; etc..

but I take him to have been implicitly concerned with the point here.)

Kreisel, as I'm understanding him, showed how formal completeness theorems bear on this question. Suppose we are antecedently convinced of (P-Sound) and (M-Comp), i.e.

Whenever $\Gamma \vdash_S B$, $\Gamma \Rightarrow B$

and Whenever $\Gamma \Rightarrow B$, $\Gamma \vDash_M B$.

A formal completeness theorem then tells us that whenever $\Gamma \vDash_M B$, $\Gamma \vdash_S B$; so it follows that the three notions $\Gamma \vdash_S B$, $\Gamma \Rightarrow B$, and $\Gamma \vDash_M B$ all coincide in extension.

Of course, this argument for their extensional equivalence (which is called the "squeezing argument") turns not just on the formal completeness theorem, but on the assumptions (P-Sound) and (M-Comp).

This Kreiselian account makes no use of the formal soundness proof. We could imagine a parallel situation, where we are antecedently convinced of (M-Sound) and (P-Comp), i.e.

Whenever $\Gamma \vDash_M B$, $\Gamma \Rightarrow B$

and Whenever $\Gamma \Rightarrow B$, $\Gamma \vdash_S B$.

Then a parallel argument using formal soundness would give you the same conclusion, that the three notions $\Gamma \vdash_S B$, $\Gamma \Rightarrow B$, and $\Gamma \vDash_M B$ all coincide in extension. But this situation has a strong air of unreality, because it's hard to imagine a situation where we are antecedently convinced of the completeness of a typical proof procedure.

Thus a Kreiselian analysis provides some explanation of the significance we attach to formal completeness proofs, but not to the significance we attach to formal soundness proofs.

Even without the completeness assumption (P-Comp) for the proof procedure, the formal soundness result plus the soundness assumption (M-Sound) for the model theory yields the soundness (P-Sound) of the proof procedure. But two considerations tend to undermine the significance of this.

First and most obviously, it won't help persuade a typical person who thinks the logic too powerful (e.g. an intuitionist is not going to be convinced that classical logic is sound by a formal soundness theorem): for such a person won't regard the model theory as genuinely sound.

Second, even an advocate of the logic in question is unlikely to find the model theoretic soundness *obvious*, except in very simple cases. Indeed, Kreisel's whole point was that it isn't obvious in the case of classical quantification theory. It isn't obvious there because it says that an argument that preserves truth *in all models* is valid; but because models have sets as their domains, they must misrepresent reality, which is too big to be a set. So preserving truth in all models doesn't obviously guarantee preserving truth in the real world, let alone in all logically possibilities. On Kreisel's analysis, we need to argue for (M-Sound), which we do by means of (P-Sound) and formal completeness. Given this, it would seem blatantly circular to use (M-Sound) to argue for (P-Sound).[34]

This is not to deny that trying to give a formal soundness proof for a proof procedure can expose technical errors in its formulation (such as forgetting to impose restrictions on the substitution of terms for variables, or forgetting to impose existence assumptions at certain points in a proof procedure for a free logic). By the same token, a successful formal soundness proof offers reassurance that one hasn't made that sort of technical error. But this is a very minimal role for soundness proofs.

Is there any prospect of proving either (P-Sound) or (M-Sound) without relying on the other?

Obviously not, if the proof is supposed to persuade adherents of other logics; but suppose the question is just whether we can provide a proof *to ourselves,* so that the proof can use the full logic for which S is a proof procedure and M a model theory.

Even on this liberal interpretation, it's obvious that no such proof is possible if we leave the genuine validity relation ⇒ as a primitive relation and don't make any assumptions about it: without putting some soundness assumption about it in, there's no getting one out.

But suppose (contrary to my argument earlier) that we were to understand validity as necessary truth-preservation? The question then is whether there's any hope of proving

> (P-Sound*) For any set Γ of sentences and any sentence *B*, if $\Gamma \vdash_S B$ and all members of Γ are true then *B* is true.

The fact is that there is no hope whatever of proving this when S contains all the logical principles we employ, *even if we're not restricted to a proof that would convince advocates of other logics, but are allowed to use the full logic codified in S in our proof*. One might think it

[34] I'll qualify this conclusion later in this section (at the end of its next-to-last paragraph).

could be done, on the basis of some sort of inductive argument, but it can't (as long as the theory in which one is doing the arguing is absolutely consistent, i.e. non-trivial). This follows from Gödel's Second Incompleteness Theorem, as I've argued elsewhere.[35]

I will not repeat that argument here, except to note that it established that *we can't even prove the special case of (P-Sound*) in which Γ is required to be empty*. This is quite significant, since despite the limitations on the connection of validity of inferences to truth-preservation, there is little doubt that it is part of the intuitive notion of validity that valid *sentences* are all *true*. (See point (B) at the end of Section 1.) Thus the inability to prove, *even using our logic*, that (P-Sound) and (M-Sound) hold *even in the restricted case where Γ is empty*, means that there is a really serious limitation in our ability to prove the soundness of our own logic.

In summary: To establish that genuine validity lines up with the model theoretical and proof theoretical notions, we have to rely on one of the soundness assumptions (P-Sound) and (M-Sound). Not only is there no hope of proving that assumption to advocates of other logics (which is hardly surprising), but *there is no hope of proving it even to ourselves, using our own logic freely*. (Assuming, again, that what S and M are proof procedures and model theories for is our full logic, not just a fragment.) For Kreiselian reasons, (P-Sound) is at least in some ways more intuitively compelling than (M-Sound), at least as regards the quantificational component of the logic. But perhaps each is compelling in a different way; and since (M-Comp) transmits whatever independent credibility it has to (P-Comp) via the formal soundness proof, this would give a bit of a philosophical role for formal soundness proofs.

As an aside, the situation is interestingly different for completeness assumptions. As I've mentioned, for standard proof procedures (P-Comp) has no antecedent claim to credence independent of the formal completeness theorem; on the other hand, there is often a reasonably compelling (I don't say airtight) argument for the completeness claim (M-Comp). I sketch it in a footnote.[36]

---

[35] I discuss it on the assumption of classical logic in Sections 2.3, 11.5 and 12.4 of Field 2008, and for various alternative logics in other places in the book. (See my response to Vann McGee in Field 2010 for some clarification.)

[36] Recall that *logical* validity is something like validity on the basis of logical form. That means that it can be valid only if all possible arguments of that form are valid. (Sentential form if it's

6. More on model-theory and proof theory. The question of soundness and completeness proofs is only part of a more general question: if as I've claimed the notion of validity is neither model-theoretic nor proof-theoretic, then why are model-theoretic and proof-theoretic analogs of the notion of validity important?

I think they are important for several reasons, and I will not try to offer an exhaustive account. One, implicit in the previous section, is that they provide a useful means for investigating the notion of direct interest, real validity. For instance, to the extent that a "squeezing argument" is available to show that real validity coincides extensionally with a certain proof-theoretic relation and a certain model-theoretic relation, then certain mathematical features of those relations (e.g. decidability or effective generability) extend to the former. Those features might be hard to establish for validity directly since it is not defined in mathematical terms.

A second and still more obvious reason is that proof theory and model theory provide useful tools for finding out what's valid and what isn't. This is especially obvious in situations where the question of which logic is correct isn't at issue; then the question of whether a

---

sentential logic in question, quantificational form if quantificational logic, and so on.) For (M-Comp) to fail, there would then have to be an argument Γ/B such that all possible arguments Γ^*/B^* of the same form are valid and yet Γ/B is not model-theoretically valid. But model-theoretic validity is usually specified in such a way that

(a) for Γ/B not to be model-theoretically valid is for there to be a model theoretic interpretation in which all members of Γ are "designated" and B is not "designated".

And model theoretic interpretations are usually defined in such a way that we can obtain from any such model theoretic interpretation a substitution function from the given language into it or an expansion of it, which "preserves logical form" and takes sentences that come out designated into well-behaved determinate truths and sentences that don't come out designated into well-behaved sentences that aren't determinate truths. (The hedge 'well-behaved' here is intended to keep you from substituting in "paradoxical" sentences.) If so, then

(b) If Γ/B (in language L) is not model-theoretically valid then there is an argument Γ^*/B^* (in L or an expansion of it) of the same form, such that all members of Γ^* are well-behaved and determinately true and B^* is well-behaved but not determinately true.

But if all members of Γ^* are well-behaved and determinately true and B^* is well-behaved but not determinately true, then presumably the argument from Γ^* to B^* isn't valid; and since it has the form of the argument Γ/B, that argument isn't valid either. So 'not model-theoretically valid' implies 'not valid'; contraposing, we get (M-Comp). To repeat, this argument is not airtight, but I think it has considerable force.

complicated inference is valid reduces to the question of whether it's valid *in the agreed-on logic*. A proof theory for a logic seems especially useful in showing what *is* valid in the logic, a model theory especially useful for showing what *isn't*. (Even in the absence of complete agreement about logic, proofs in a natural proof theory can be highly compelling: they persuade us to constrain our degrees of belief in accordance with them. Similarly, counter-models in a natural model-theory can persuade us not to constrain our degrees of belief in a way that would automatically rule them out. This is a main reason why *natural* proof and model theories are better than mere algebraic tools that yield the same verdicts on validity.)

My third reason for the importance of model theory and proof theory, and the one I most want to emphasize, is that they provide a useful means of communication between adherents of different logics. And without communication, there is no chance of intelligent debate between adherents of different logics.

The reason they aid communication is that many adherents of non-classical logics think that classical logic is "in effect valid" throughout much of mathematics—enough mathematics to do proof theory and usually enough to do model theory as well. For instance, an adherent of quantum logic is likely to think that mathematical objects such as proofs and models can't undergo superpositions, and that for objects that can't, the distributive laws are correct: so the distributive laws should be accepted as (not strictly logical) laws *within mathematics*. And quantum reasoning *from those laws* is in effect the same as classical reasoning. The same holds for many other non-classical positions: for instance, if excluded middle is problematic only for sentences containing 'true' and related terms, then since such terms aren't part of proof theory or model theory,[37] excluded middle is in effective valid within proof theory and model theory.[38] Given this, proof theory and model theory provide a ground that is neutral between many adherents of different logics.

[37] Obviously 'true in model M' occurs in model theory, but it is to be distinguished from 'true': e.g. 'true in M' is set theoretically definable, but 'true (in set theory)' isn't. See the discussion that starts in the paragraph after next.

[38] And if excluded middle is problematic only there or when vagueness is relevant, then *to the extent that* there is no vagueness in the mathematics employed in proof theory and model theory, excluded middle is effectively valid within those disciplines.

This marks an important contrast between model theory and proof theory on the one hand and the alleged definition of validity in terms of necessary truth preservation. I've agreed that that alleged definition gives the right results in most instances, and at the moment I'm not worried about the cases where it fails; so for present purposes I could even concede that it is a correct definition. But even if correct—and even if it better captures the meaning of 'valid' than either proof-theoretic accounts or model-theoretic accounts do—it would be useless for certain purposes. For instance, it would be useless to employ it *in trying to explain one's logic to an adherent of a different logic*. The reason is that the adherents of the different logics disagree about which arguments preserve truth by logical necessity, so if I tell you that the valid inferences are (plus or minus a bit) those that preserve truth by logical necessity, that will convey very little about my logic. Whereas if I provide you with a model theory or proof theory for my logic, you will almost certainly reason from this information in the same way that I do, so there is real communication.[39]

This points up a fundamental difference between the concept of truth and the model-theoretic property of "designatedness" or "truth in a model". A typical model theory M for a non-classical logic will be one in which certain classical arguments (say, excluded middle) don't preserve the property of being "designated in M-models", where that property is specified in purely mathematical terms. There won't be serious disagreement as to which ones do and which ones don't: that's a purely mathematical question. The question of which ones preserve truth has an entirely different character, for two reasons.

First, even the advocate of the logic whose model theory is M usually won't identify being true with being designated in some particular M-model. For even one who doubts the general applicability of excluded middle is likely to hold that it applies to precise properties such as being designated in a given M-model. But such a person will presumably think that it doesn't apply to the property of truth: if it's wrong to accept a given instance of excluded middle '*A* or not-*A*', then presumably it's equally wrong to accept 'either ⟨*A*⟩ is true or ⟨*A*⟩ is not true'. Since the advocate of the nonclassical logic thinks that designatedness obeys excluded middle and truth doesn't, she can't identify truth with designatedness. (Similarly for laws other than excluded middle: a logic

[39] In principle, one party could also tell the other party his or her constraints on degrees of belief; that too is unlikely to be understood in the wrong way. But giving a model theory or proof theory is easier.

that allows assertions of ‘both *A* and not-*A*’ should presumably equally allow corresponding assertions ‘⟨*A*⟩ is both true and not true’, which prevents truth from being identified with any property definable in a model theory for that logic if the model theory is stated in classical logic.)[40]

This first reason is *slightly* controversial, in that it depends on a view that isn’t universally accepted about how we should use the term ‘true’ if we advocate a non-classical logic. The second reason is quite independent of this: it is that even if we could somehow identify being true with being designated in an appropriate model *of the correct logic*, still the presence of the word ‘correct’ would preclude using this to explain one’s logic to someone else. (If I say to you, “The arguments I take to be correct are those that preserve designatedness in models of the correct logic”, that won’t give you a clue as to what logic I accept.) Model theory, and proof theory too, are useful in explaining logic since they provide specifications of which arguments are valid that can be applied in the same way by adherents of different views as to what’s correct.

There are some limits here: a person with a logic so weird that he reasoned in a very different way than the rest of us do *from premises within mathematics* couldn’t really understand our proof-theoretic or model-theoretic explanations of our logic, for he would reason from what we said in very different ways than we do. But cases of that sort are mostly a philosopher’s fiction: for the kind of disputes about logic that people actually have, proof theory and model theory provide an extremely useful tool of clarification.[41]

Hartry Field
New York University

References

[40] A related point is that even a classical logician won’t identify truth with designatedness in models of quantificational logic: this is because of the fact (used to a different purpose above in discussing Kreisel) that there is no model big enough to contain everything, so no model corresponding to the actual world, so that there is no model such that being designated in (or true in) that model corresponds to being true.

[41] Thanks to Paul Boghossian, Daniel Boyd, Josh Dever, Sinan Dogramaci, Kit Fine, Paul Horwich, Jeff Russell and an anonymous reviewer for some criticisms that have led to improvements.

Adams, Ernest 1975, *The Logic of Conditionals*. Reidel Publishers.

Blackburn, Simon 1980, "Opinions and Chances". In D. H. Mellor, *Prospects for Pragmatism* (Cambridge University).

Edgington, Dorothy 1995, "On Conditionals", *Mind* 104: 235-329.

Field, Hartry 2001, *Truth and the Absence of Fact*, Oxford University Press.

__________ 2003, "No Fact of the Matter", *Australasian Journal of Philosophy* 81: 457-80.

__________ 2008, *Saving Truth from Paradox*, Oxford University Press.

__________ 2009, "What is the Normative Role of Logic?", *Proceedings of the Aristotelian Society*, Supplement 83: 251-68.

__________ 2010, "Reply to Vann McGee", *Philosophical Studies* 147: 457-60.

Fine, Kit 1975, "Vagueness Truth and Logic", *Synthese* 30: 265-300.

Hyde, Dominic 1997, "From Heaps and Gaps to Heaps of Gluts", *Mind* 106: 641-60.

Jeffrey, Richard 1965, *The Logic of Decision*, McGraw-Hill.

Kreisel, Georg 1967, "Informal Rigor and Completeness Proofs". In I. Lakatos, ed., *Problems in the Philosophy of Mathematics* (North-Holland).

Lewis, David 1980, "A Subjectivist's Guide to Objective Chance". In R. Jeffrey, ed., *Studies in Inductive Logic and Probability, vol. 2* (University of California).

MacFarlane, John 2010, "Fuzzy Epistemicism". In R. Dietz and S. Moruzzi, *Cuts and Clouds: Vagueness, its Nature, and its Logic* (Oxford University).

__________ (unpublished), "In What Sense (If Any) Is Logic Normative for Thought?". Delivered at the American Philosophical Association Central Division conference, 2004.

McGee, Vann 1985, "A Counterexample to Modus Ponens", *Journal of Philosophy* 82: 462-71.

Restall, Greg 2005, "Multiple Conclusions". In P. Hajek, L. Valdez-Villanueva and D. Westerstahl (eds.), *Proceedings of the 12th International Congress of Logic, Methodology* *and Philosophy of Science* (King's College).

Schiffer, Stephen 2003, *The Things We Mean*. Oxford University Press.

Shafer, Glen 1976, *A Mathematical Theory of Evidence*. Princeton University Press.

Skyrms, Brian 1984, *Pragmatics and Empiricism*. Yale University Press.

Tarski, Alfred 1936, "On the Concept of Logical Consequence". Translation by J. H. Woodger in Tarski, *Logic, Semantics and Metamathematics*, Clarendon Press, Oxford, 1956.