

# Why Is a Valid Inference a Good Inference?

SINAN DOGRAMACI

*The University of Texas at Austin*

True beliefs and truth-preserving inferences are, in some sense, good beliefs and good inferences. When an inference is *valid* though, it is not merely truth-preserving, but truth-preserving *in all cases*. This motivates my question: I consider a Modus Ponens inference, and I ask what its validity in particular contributes to the explanation of why the inference is, in any sense, a good inference. I consider the question under three different definitions of ‘case’, and hence of ‘validity’: (i) the orthodox definition given in terms of interpretations or models, (ii) a metaphysical definition given in terms of possible worlds, and (iii) a substitutional definition defended by Quine. I argue that the orthodox notion is poorly suited to explain what’s good about a Modus Ponens inference. I argue that there is something good that is explained by a certain kind of truth across possible worlds, but the explanation is not provided by metaphysical validity in particular; nothing of value is explained by truth across *all* possible worlds. Finally, I argue that the substitutional notion of validity allows us to correctly explain what is good about a valid inference.

## 1. Introduction

### 1.1. *The Question*

Open an introductory logic textbook to page 1, and you are likely to read a claim to the effect that logic is the study of “good” reasoning.<sup>1</sup> When you read on, though, you find that all talk of *good* is quietly dropped, replaced only with talk of *validity*.

---

<sup>1</sup> See Barwise and Etchemendy (1999: p.1), Bergmann, Moor and Nelson (1980/2008: p.1), Bonevac (1987/2003: p.1), Goldfarb (2003: p.xiii), Prior (1962: p.1), Restall (2006: p.1), Sainsbury (1991/2001: pp.1–2), Salmon (1963/73: p.1), Teller (1989: pp.1–2), and Tomassi (1999: p.2).

Frege’s *Basic Laws of Arithmetic* begins with similar suggestions that logic is normative for thought. See Frege (1893/1997: p.xv/p.202). Also see his ‘Thought’, Frege (1918/1997: the opening paragraph).

Russell and Whitehead, in *Principia Mathematica*, are more terse and less explicit than Frege. They say they identify principles of logical implication with ‘the principles by which conclusions are inferred from premises’. See Russell and Whitehead (1910: first two paragraphs of *Section A*).

Let's look at these two notions, *good* and *valid*, using a concrete example. Suppose you adopt a belief that NY won on the basis of rational beliefs that LA lost, and that NY won if LA lost. Suppose also that you continue to rationally believe those two premises, and suppose you lack any independent reasons to doubt NY won. We can at least conceptually distinguish, in this Modus Ponens inference, the following two features. First, the logical feature: the inference is *valid*, which I will understand to mean the transition from premises to conclusion is truth-preserving “necessarily” or “in all cases”. (For treatments of this as the core notion of validity, see, e.g., Jeffrey (1967/2006: p.1) and Beall and Restall (2006: pp.23, 29). As I'll discuss below, those scare-quoted terms can be understood in importantly different ways.) Second, there is a normative feature: the inference is, in some way, *good* or *valuable*. We praise the inference by saying that you acquired a rational belief, and we say this because the inference is indeed praiseworthy, so there must be at least something *worth caring about*.

In this paper, I want to examine the relationship between these logical and normative features, and I will focus on Modus Ponens. In the described Modus Ponens inference, there indeed is something good, something valuable, something worth caring about and praising. But, what is that good thing, and what does validity have to do with it?

The question of this paper is this: how, if at all, does a Modus Ponens inference's *validity in particular* help explain why it is, in any sense, a *good* inference? If it is true that the inference is, in any sense, good because it is valid, then how is that so?

## 1.2. Why the Question Is Interesting

I stated my main question using ‘good’, asking if validity explains why *any sense* of that term applies to a Modus Ponens inference. I mean ‘good’ as an ordinary and generic normative term, one that I take to be interchangeable with other ordinary normative labels like ‘valuable’ or even ‘is worth caring about’ or ‘is something we should care about’. (Having the other terms makes some locutions neater, e.g. I can use ‘value’ instead of the clumsier ‘goodness’ or ‘what’s good’.) But, you may already be asking: in what sense of ‘good’ could I be imagining that an inference, such as the Modus Ponens inference that NY won, qualifies as *good*?

Well, one thing that's uncontroversially good about it—maybe the *only* uncontroversially good thing—is that the reasoning actually preserves truth: if the believed premises are true, then you'll add a true belief. Why is truth good? You are welcome to take it as a basic assumption of this paper that a belief's being true is a good thing. I myself am attracted to the following simple account. Because of the way in which a person's actions are normally based on her beliefs and desires, believing the truth helps her satisfy

her desires. The general idea is that you act in ways that *would* tend to satisfy your desires *were* your beliefs true. So, if what you believe *is* true, your desires *will* tend to be satisfied. (See, e.g., Stalnaker (1984: p.15).) A slightly more specific suggestion adds that the directly useful true beliefs, the ones your actions are directly based on, are those that concern which actions will satisfy your desires. All other true beliefs are indirectly useful, because the directly useful beliefs might need to be acquired inferentially from any other beliefs, so you want all your beliefs to be true and you want your reasoning to be truth-preserving. (See Horwich (1990/98: § 3.11) and Goldman (1992: p.164).)

And it's not that truth-preservation is good only when the inference's premises are all true (and thus the conclusion is true too). No, truth-preservation is valuable even when a premise is false. (As is usual among philosophers and logicians, I count an inference as truth-preserving just in case it doesn't have all true premises and a false conclusion, that is, just in case its corresponding material conditional is true.) Truth-preservation is valuable even when a premise is false, because, if your inferences are truth-preserving, then you are especially well-positioned to *discover* any false premises. You are well-positioned because, while you may not have initially recognized the falsity of the premise, you may eventually recognize that one of your inferred conclusions is false, and then, so long as your inference was (or your chain of inferences all were) truth-preserving, you're now in a position to infer what will be the *true* conclusion that one of your original premises must have been false, and its negation true. So, quite generally, true belief is good, and truth-preserving inference is good.

What's really puzzling me, though, is that validity gives us much, much more than truth-preservation. Validity gives us truth-preservation throughout a huge further range, the range of "all cases". Why should we care about that? Does validity, or something to do with validity, add any *further contribution*, lend any *further help*, to explaining what's good about an inference? Compare: if I conjoined the fact that my inference is actually truth-preserving to an endless list of facts about, say, which numbers are multiples of the number of seconds it took me to perform my inference, that would not help explain what's good about the inference. So, what's helpful about adding to actual truth-preservation all these facts about truth-preservation in other cases? That is what motivates my interest in asking my question. How, if at all, does my inferences' being not merely truth-preserving but valid help explain why such reasoning is, in any way, good? I see the value of truth. I don't yet see the special value of truth in all cases.

### *1.3. Some Preliminaries: Explanation, Inference, Rationality, Harman*

A few preliminaries may help prevent some misunderstandings.

One mistake that's easily made is to think my question can be answered as follows: an inference's validity explains why it's truth-preserving because *the only way we can know* that it's truth-preserving is by inferring that from its validity. And, the thought goes on, if the validity thus explains the truth-preservation, then, by some transitivity principle for explanation, the validity explains why the inference is good.

This thought is simply mistaken. Even if we do, or even if we must, *infer* one fact, *B*, from another, *A*, this tells us nothing about how *A* could explain why *B*. In IBE (inference to the best explanation), we infer an explanation from some data, but of course the data does not explain the explanation (nor does it, by transitivity, explain itself). Furthermore, I deny that our knowledge of truth-preservation must be inferred from knowledge of validity. In formal textbook proofs of validity (so-called soundness proofs), the inference is the other way around: we prove validity by first proving truth-preservation for an arbitrary case and then generalizing. My own view is that we have independent knowledge of these two things, validity and truth-preservation in an arbitrary case; neither need be inferred from the other; see Dogramaci (2010).

A related mistake is to think that validity explains truth-preservation (and, by transitivity, an inference's value) by the principle that a generalization explains its instances. We cannot just place an inference's truth-preservation under a generalization and consider it an explanation. I gave an example of an obviously non-explanatory such generalization above: this inference is truth-preserving, and the number of seconds it took me to perform it was a factor of 4, of 8, of 12, and so on. In this paper, I am going to examine several conceptions of validity that do make an inference's truth-preservation an instance of a generalization (truth-preservation in all cases), but I'll argue that just one of these (just one sense of 'case') best explains what's good about the inference.

Another avoidable misunderstanding concerns the role of rationality in answering my question. I said I want to know what validity contributes to explaining why *any sense* of 'good' or 'valuable' applies to the inference that NY won. And I said I'm taking it that truth and truth-preservation are clearly good things. What about rationality then? (Or justification, or reasonableness; I don't distinguish these.) We praise the inferred belief that NY won, calling it a rational belief. So, could the answer to my main question just be: (i) rationality, like truth, is a valuable thing, and (ii) the rationality of the conclusion is explained by the inference's validity?

This answer fails to resolve my puzzlement. I see problems with each of the claims here, (i) and (ii).

I find claim (i) dissatisfying. The claim made by (i) is about *epistemic* rationality; it's the claim that epistemic rationality is valuable. And, I cannot immediately see what is valuable about epistemic rationality *independently*

*of its connection with truth.* The source of my puzzlement about validity was about how validity, i.e. truth in all cases, contributes anything valuable beyond mere actual truth. It's hardly less puzzling why an inference or belief's *epistemic rationality* contributes anything good or valuable *beyond* any value that's accounted for by truth. So, I'm not happy to rest content with just saying the rationality of the valid inference that NY won is what's valuable about it. The answer I'll eventually offer to this paper's main question won't mention rationality (as such).

To be clear, in my running example of the Modus Ponens inference that NY won, the inference is not only valid but rational as well. I am stipulating that the reasoner draws her conclusion from rationally maintained belief in the premises, and she has no independent reason to doubt her inferred conclusion. This seems to me a sufficient condition for the conclusion's rationality. Some authors have proposed including the following in a more general sufficient condition for a valid inference to be rational: they propose the validity of the inference be *realized*, or *apprehended*, or *obvious to the reasoner*. (See Sainsbury (2002), MacFarlane (ms/2004), and Field (2009b).) While I don't think such a higher-order awareness condition adds anything useful in the case of Modus Ponens (for reasons found in Carroll (1895), Boghossian (2003), and Dogramaci (2013)), you are welcome to add this feature to the example if you want. I'll point out below that the arguments I give are unaffected by whether or not it's realized by, apprehended by, or obvious to the reasoner that her conclusion follows from her premises.

I also see a problem with (ii) above, the claim that the rationality of the conclusion that NY won is explained by the inference's validity. As Harman (1986: chapter 2) observes, remote consequences may not be rationally inferred. That is, although it may be rational to believe the consequence that NY won on the basis of the above two entailing premises, these premises also validly entail all sorts of remote, obscure logical theorems and consequences which it would not be similarly rational to believe on only this basis. Thus, we cannot explain a valid inference's value by pointing to its rationality, for it might not even be rational. If an inference's validity really does contribute something of its own to explaining why it is a good inference, then the explanation must be something else.

The remoteness problem supports the main thesis of Harman (1986: chapter 2), namely that logic is not, as he puts it, *specialy relevant* to reasoning. This is because, if you address the remoteness problem in the apparently only possible way, by imposing one of Sainsbury, MacFarlane or Field's higher-order awareness requirements, then you make logical consequence just as relevant to rational reasoning as any kind of *non-logical* consequence. If you are aware that some claim follows as a matter of apriori, analytic, or even aposteriori nomological consequences, you are just as rational in inferring it as you would be if it were a matter of logical consequence. Logical consequence, validity, still is not

*especially* relevant to reasoning. (This is also noted in Harman (2009) and Boghossian and Rosen (ms/2004).) I am largely in agreement with Harman on all this. But later on I will qualify the extent of my agreement; at the end of sub-section 4.3 I'll say more about how I view the relations among rationality, validity, and validity's contribution to what's good about an inference.

#### 1.4. Plan for the Remainder of the Paper

The plan is to consider three different ways of defining 'case', and hence 'valid', and examine whether the resulting notion could help explain the value of a Modus Ponens inference, such as the inference that NY won. Section 2 considers the orthodox understanding of validity, one defined in terms of interpretations or models. There, I argue that the orthodox notion of validity doesn't help explain the value of the inference. Section 3 turns to a metaphysical understanding of validity, one that defines cases as possible worlds. While agreeing that there is *a* fact about truth preservation in possible worlds that does explain something good about the inference, I argue that metaphysical validity, truth preservation in *all* worlds, doesn't help explain what's good about it. Section 4 turns to an understanding of validity that Quine advocated, one that defines cases in terms of substitutions of terms with fixed interpretations. I argue that we must make use of this last notion if validity is to help explain the value of a Modus Ponens inference.

## 2. Interpretational Validity

### 2.1. Characterizing Interpretational Validity

Validity is truth-preservation necessarily, that is, in all cases. We get different understandings of validity by understanding 'case' differently.<sup>2</sup> Let's first consider that orthodox understanding of 'case', and likewise of 'necessarily' and 'valid', found in standard textbooks on logic, which we can call *interpretational validity*.<sup>3</sup>

---

<sup>2</sup> The methodology here is thus similar to that of Beall and Restall (2006).

<sup>3</sup> Etchemendy (1990) contrasts what he calls 'interpretational semantics' and 'representational semantics', which correspond to the notions of validity examined in this section and the next, respectively. Etchemendy is well known as a critic of interpretational validity as an analysis of genuine logical validity. He should not be misunderstood, though, as recommending the use of representational semantics, or what I'll call metaphysical validity in the next section, to provide the analysis of logical validity. He says: '[I]t would clearly be wrong to view representational semantics as giving us an adequate analysis of the notion of logical truth. . . . The value of representational semantics does not lie in an analysis of the notions of logical truth and logical consequence, or in the analysis of necessary or analytic truth.' As he makes somewhat clearer in a later follow-up paper, Etchemendy (2008), he is skeptical of the *reductive* approach to understanding logical validity.

On this approach, we define a case to be any uniform assignment of truth-conditional contents just to the non-logical terms. The most common examples of truth-conditional contents are referents (assigned to names) and extensions (assigned to predicates). Call such a case an *interpretation*.

A few clarificatory comments. We assign contents *uniformly* in the sense that reoccurring non-logical terms, e.g. the name ‘NY’, receive the same assignment, e.g. the New York Yankees (or the Chicago Cubs, or my mother, or the planet Venus, or the number 534, or whatever). In our NY/LA example, the only logical term is ‘if’. For our purposes throughout this paper, we can presuppose a reasonably standard enumeration of the logical terms. We can remain agnostic about what explains why a certain enumeration is the correct one, if there is a correct one in any interesting sense. (In particular, the arguments of this paper are compatible with any of the accounts of the logical constants canvassed in MacFarlane (2009).) In first-order predicate logic, the definition of ‘case’ is slightly more complicated than the simpler definition I’ve given here: since quantifiers are standard logical terms, cases in predicate logic consist not only of reinterpretations of truth-conditional content, but also of varying *domains* for the quantifiers to range over. This gives the notion standardly called a *model*, standardly credited to Tarski (1936/1983).<sup>4</sup> However, domains play an idle role in our NY/LA example since that example involves no quantifier, so we can stick with the simpler definition of a case as just an interpretation, at least for now.<sup>5</sup>

On this interpretational understanding of case, then, we have it that an inference is *interpretationally valid* if and only if there is no interpretation on which the premises are all true but the conclusion is false.

What sort of things are the premises and conclusions of an *inference*: are they the reasoner’s beliefs, or are they the contents of those beliefs? The natural answer here is that they are the beliefs, since we are talking about assigning different contents to these things—and the idea of assigning contents to contents seems awkward at best. Interpretational validity thus presupposes that beliefs have something like linguistic structure, something that allows the reinterpretation of *terms* that occur in beliefs, things like names and predicates. The simplest way to understand this is to suppose that a belief is a sentence in mentalese, the language of thought. Even if you think a belief, say, the belief that NY won, lacks any intrinsic structure, you

---

<sup>4</sup> Though, see Etchemendy (1988) for a famous argument that Tarski did not introduce the standard, modern notion of a model with a varying domain. See Mancosu (2010) for a recent assessment of the ensuing debate.

<sup>5</sup> Technical detail: one role of domains in first-order logic, a role they have even for a sentence lacking quantifiers, is to provide the set from which referents and extensions are selected. So, either we can go ahead and give each interpretation a domain, or we can just select referents and extensions from wherever the members of model-theoretic domains are selected from in the first place.

may still attribute it the requisite structure by courtesy of a certain structured sentence, perhaps the sentence (that would be) used by the subject to express the belief, or perhaps the sentence (or that-clause) used by an evaluator attributing validity, presumably ‘NY won’. I will assume that one or another way of attributing the requisite structure to beliefs is acceptable. Just for convenience though, I’ll often talk as if we think in mentalese.

## 2.2. Does Interpretational Validity Explain Why an Inference is Good?

Well, even though it is the orthodox definition of validity, and even though modern logic’s founders and textbook authors tell us that validity has normative import, it is not easy to see how interpretational validity helps explain what is so good about my inference. The problem is not very complicated. The problem is simply that it is mysterious what *alternative interpretations* of my beliefs should have to do with any normative features of my concluding what I actually concluded on the basis of what I actually believed.

To see the problem a little more vividly, look at the simple NY/LA example. We suppose I start with two mentalese sentences, ‘LA lost’ and ‘NY won if LA lost’, and, given the use-properties these sentences actually have, what they mean is that LA lost and that NY won if LA lost, respectively. I then add a new sentence, ‘NY won’, and what this sentence means, given its use, is that NY won. The inference is indeed interpretationally valid. But does that help explain why it is in any way good or valuable? It is hard to see why interpretational validity is worth caring about at all. Why should it be good that, if my mentalese sentences were assigned certain alternative meanings, ones *completely different from the thoughts I actually had*, I would still not draw a false conclusion from true premises? It’s not apparent how that fact could help to explain what is good about the inference. Interpretationally valid reasoning is insured against a kind of error, but a kind of error that it is not clearly worthwhile to insure one’s reasoning against. Was there any *doubt* about what I meant? Unless I am uncertain about the actual interpretation of my belief states—and it is hard to see why I ever would be—there is little value in insuring that my reasoning preserves truth even if my thoughts are about topics *I know they are not about*. Why is it worth caring that, if my sentence ‘LA lost’ had expressed the thought that LA won, or even that pigs fly, then I would still not have inferred a falsehood from truths?

The value of interpretational validity is no clearer if we consider the reasoner third-personally. Suppose Smith inferred that NY won on the basis of beliefs that LA lost and NY won if LA lost. There is something we find valuable, something worth praising, about Smith’s reasoning. What is that



thing? Is it that Smith's reasoning would have preserved truth if her belief states had represented different facts? Let it be true that, if they *had*, then Smith's reasoning *would have* still been valuable, still been worth praising. That doesn't explain what is good about what Smith *did*, that is, what Smith *actually did*. Except perhaps in some extremely odd circumstances, we do not evaluate Smith's reasoning, and find it to be praiseworthy, while under any confusion or doubt about what she was thinking about: she was thinking about NY and LA, winning and losing.

Finally, also note that, if the rationality of valid reasoning does require an apprehension or obviousness or realization condition, the value of interpretational validity only becomes more obscure. It's hard to imagine how subjects who apprehend, realize, or find obvious the validity of their inferences could be in any doubt about what interpretation their beliefs actually have.

So, interpretational validity seems to not have any explanatory relevance to what makes reasoning good. We don't satisfyingly explain why a valid inference is a good inference by observing that it is interpretationally valid.

The problem here reminds me a bit of Kripke's famous critical remark about counterpart theory: Humphrey cares that he could have won, but 'Humphrey could not care less whether someone else [his counterpart], no matter how much resembling him, would have been victorious in another possible world.' Kripke (1972/80: p.45). But, in fact, things look far worse for interpretational validity than for counterpart theory: we pre-theoretically care about what could have been, no matter what we post-theoretically identify that with; by contrast, nobody pre-theoretically cares about interpretational validity.

The problem is also reminiscent of a famous debate between Frege and Hilbert.<sup>6</sup> Hilbert's view, a precursor to the orthodoxy that followed Tarski's treatment, was that logical relations hold among uninterpreted sentences, depending on the existence of the appropriate interpretations. Frege's contrasting view was that logical relations hold among contents (or *thoughts*, as he called them), not uninterpreted sentences.<sup>7</sup> While it's not well understood (at least by me) exactly what Frege understood validity to consist in, I've been expressing sympathy in this section for Frege's lack of interest in Hilbert's interpretational approach. If there is some good reason that the interpretational notion became orthodoxy, I cannot see that it is due to its ability to explain what's good about an interpretationally valid inference.

---

<sup>6</sup> There is disagreement over how to understand Frege's views here. I am reporting the reading defended by Blanchette (1996), among others. See Blanchette (2012) for more overview. The key primary source is Frege (1906/84: especially sections II and III).

<sup>7</sup> This is explicitly claimed by Blanchette (1996), Burge (1998: p.322), and Tappenden (2000: p.282). For additional relevant support, see Stanley (1996: pp.60–3) and Heck (2007: p. 48, also pp.43–6).

Perhaps some will argue interpretational validity's explanatory power is due to features of interpretational validity that, perhaps non-obviously, turn out to be by-products of the property described in its definition (as truth in all interpretations). Most notably, one could argue that standard model-theoretic validity, an interpretational notion, turns out to also be a species of *necessity*, in an interestingly strong modal sense (as argued, for example, by Sher, Hanson and Shapiro).<sup>8</sup> I don't think this will help with our question, though. A general problem with this strategy is that the property of being interpretationally valid would then not be part of what explains why a valid inference is a good inference; rather, only the other by-product property would figure in the explanation. (I will argue further for this point about the structure of explanation below, in sub-section 4.5.) And anyway, at least for metaphysical necessity as the suggested by-product property, there is a more particular problem: even if validity is defined in terms of metaphysical modality, it is not fit to explain why a valid inference is a good inference, as I'll argue next.

### 3. Metaphysical Validity

#### 3.1. Characterizing Metaphysical Validity

So, let's turn away, at least for now, from the way validity is explained in logic textbooks, and turn to another way of understanding the idea of truth in all cases. Let's now consider cases as possible *worlds*. Rather than considering possible ways the terms in our given inference could be reinterpreted, we'll now consider possible ways the world could be and we'll consider what effects this has on the truth-values in our inference *given its actual interpretation, kept fixed*.<sup>9</sup>

---

<sup>8</sup> Sher defends a model-theoretic view of consequence that she describes as intuitively capturing a species of necessity. Her view is developed in several works, including Sher (1991) and Sher (1996: see especially pp.668, 674).

Hanson and Shapiro both propose to define validity in an explicitly conjunctive way, defining it as arguments that both preserve truth in all interpretations and in all possible worlds. Shapiro (1998: p.148) proposes this definition (calling it a 'conglomeration' or a 'blend'): 'Φ is a logical consequence of Γ if Φ holds in all possibilities under every interpretation of the non-logical terminology in which Γ holds.' See Hanson (1997: pp.379, 390) for almost the same view. Hanson and Shapiro both then go on to argue that their proposed definitions turn out to coincide with the standard, model-theoretic definition (at least for first-order logic).

Beyond critiquing the individual interpretational and metaphysical elements of such a conjunctive definition, I won't separately discuss whether a conjunctive notion might succeed in explaining why a valid inference is a good inference, where its individual elements fail. It will be clear that resorting to such a conjunction won't help.

<sup>9</sup> As mentioned above (see footnote 3), Etchemendy gives extensive discussion to the similar notion of *representational semantics*, but he does not believe this provides an analysis of logical consequence.

The simple and familiar notion of a possible world is the sort popularized by Kripke (1972/80) and Lewis (1986). The notion is non-linguistic, and is meant to be present in our ordinary understanding of necessity and possibility. It is the sense of possibility we express when we say it could not have been that LA lost, and NY won if LA lost, and NY did *not* win. Call the associated notion of necessary truth-preservation *metaphysical validity*. So, let's now ask: could the metaphysical validity of an inference, that is, could the fact that there is no possible world in which the inference fails to preserve truth, help explain why it is good or valuable?

You might worry metaphysical validity cannot explain the value of valid reasoning, because it is sometimes not apriori whether an inference is metaphysically valid. (Take an inference that concludes that water is H<sub>2</sub>O.) This might tempt you to shift attention to a notion of validity defined using *epistemically* possible worlds, roughly, ways the world might turn out to be. One issue here is that, typically, the full definition of an epistemic possibility uses notions that are themselves normative, and since we are trying to ask what explains the value of validity, there is some risk of ending up with a shallow explanation, e.g. if epistemic possibilities are defined, as they are in Chalmers (2012), as whatever *good*, or *rational*, *apriori reasoning* cannot rule out. Another alternative you might be tempted to reach for, again perhaps if the existence of aposteriori entailments worries you, is a “metaphysical” definition of validity suggested by Fine (1994: pp.9–10) and Shalkowski (2004: section V), one that would seem to make validity always apriori: as Fine put it, ‘logical necessities can be taken to be the propositions which are true in virtue of the nature of the logical concepts’.

But, I think there is no reason to be worried by aposteriori metaphysical entailments in the first place. The phenomenon of aposteriori entailment just raises another version of Harman's remoteness worry: for certain consequences, it will be obscure to a reasoner that they are entailed by known premises. So, we can leave the definition in its simple form, truth preservation in all possible worlds, and those who want to may impose the apprehension/realization/obviousness condition on when validity is relevant to reasoning—in fact, aposteriori consequence was the issue that initially motivated MacFarlane (ms/2004) to impose an apprehension condition. (Speaking for myself, though, I still don't see how apprehension, realization, or obviousness will help us explain what's good about the NY/LA inference, which is apriori metaphysically valid anyway.)

I should explicitly acknowledge that it is unclear whether metaphysical validity is a species of *formal* validity. But, I started out by saying I will understand validity to just mean truth-preservation necessarily, or in all cases, and metaphysical validity is clearly a species of this sense of validity. For those who believe formality is important, Beall and Restall (2006: section 4.2.2) argue that metaphysical validity qualifies as formal in some

senses (they apply various notions of formality distinguished by MacFarlane (2000)), and another option is, again, to attempt to develop Fine or Shalkowski's notion.

### 3.2. *An Inference's Reliability Does Explain Something Good about It*

Ultimately, we want to know if, and how, the metaphysical validity of the inference that NY won might help explain why it's a good inference. In this sub-section, I'll argue that there is *a* fact concerning truth-preservation in other possible worlds that does help to explain one important part of why the inference is good. To the extent that our puzzlement is just about the value of any kind of truth-preservation beyond mere actual truth-preservation, this constitutes progress. However, the next sub-section will argue that we've made limited progress on our main question about the contribution of *validity*. There, I'll argue against the claim that an inference has any special value because it is *metaphysically valid*.

It's a natural thought that the *reliability* of the inference that NY won helps explain a way in which the inference is valuable. But, just on its own, the observation that the inference is reliable won't resolve any puzzlement over why there is anything valuable about non-actual truth-preservation. For, reliability is naturally understood to involve truth-preservation in non-actual though nearby worlds. How, though, does non-actual truth contribute anything valuable? It's important to appreciate that something other than mere reliability needs to be mentioned to give the explanation.

Craig (1990: pp.19–20) provides the key answer. He rightly observes that, when we don't know which future is actual, it is rational, in the *practical* sense of 'rational', to make plans or preparations that will serve well in *multiple* possible futures, at least some of which have to be non-actual (since there's only one actual future). I don't know whether or not it will rain tomorrow, so it will be useful to me if I can form beliefs (on topics other than tomorrow's weather) that will be true in either possibility, whichever turns out to be actual. I can do this by drawing inferences that preserve truth in both possibilities, something metaphysically valid inferences of course do. That is how the truth-preservation of our reasoning in some non-actual worlds is something worth wanting, something valuable. The value is practical.

(You might think that this sort of truth-preservation is only useful, or perhaps is more useful, when it is apprehended. I'm unsure, but again, I just don't want to argue over the apprehension condition.)

Alongside Craig's explanation, there may be a second, distinct proposal about how truth-preservation in other worlds contributes to the value of an inference. This proposal is premised on the thought that suppositional reasoning generates knowledge of a counterfactual conditional by *the*

*reasoning's preserving truth in the nearest world where the counterfactual's antecedent is true.* This proposal's premise will be controversial. (The premise is not simply Lewis (1973) and Stalnaker (1968)'s theory of a counterfactual's truth-conditions; the premise is a view about how suppositional reasoning generates knowledge of a counterfactual, namely by virtue of preserving truth in the nearest world where the antecedent is true. Such a view is not plausible for knowledge of any conditional whose truth-condition is actual truth-preservation; plausibly, to know such a conditional, one's suppositional reasoning need only actually preserve truth.) In any case, this proposal, if right, would show how certain reasoning's truth-preservation in non-actual worlds would be valuable. Suppose a scientist wants to know a conditional linking a hypothesis to its predicted observations. The suggestion is that such a conditional can be learned by engaging in suppositional reasoning that preserves truth in the nearest worlds where the hypothesis is true, even if those worlds are non-actual. It's valuable to the scientist to know the conditional because she can then refute hypotheses whose predictions, she observes, do not actually obtain. Since this second proposal about the value of reliability creates controversy whereas Craig's should not, I'll restrict attention to Craig's proposal from here on.

### *3.3. Does an Inference's Preserving Truth in Every Last Possible World Help Explain What's Good about It?*

Craig's point about planning explains why there is some value, practical value, in reasoning that will remain true, not just in the actual future but certain others, specifically those we do not know to be non-actual. However, this is an explanation of the value of the inference that NY won which concerns only one *part* of that inference's metaphysical validity. The metaphysical validity of the inference that NY won involves truth-preservation in a vast further range of known-to-be-non-actual worlds, and Craig's point leaves this large extra measure of truth-preservation explanatorily redundant. If I infer that NY won, does my inference possess some value that is specifically explained by its remaining truth-preserving in *all* possible worlds, without exception? The answer is, no: there's no added value to my inference's preserving truth in the countless possibilities *I do not need to plan for*.

Let me make clear what I am, and am not, claiming. My aim, here in this sub-section, is not to argue against the claim, already endorsed, that something valuable about my inference that NY won is explained by its preserving truth not only actually, but in a certain range of other possible worlds too. What I am arguing in this sub-section is that we haven't got any explanation of anything good about the inference where the explanation should include the claim that the inference is metaphysically valid. There is nothing good that is explained by metaphysical validity *in particular*.

To see why this is so, consider some fairly uncontroversial general points about the nature of planning and rational action. When I plan what to do, I am guided by my actual knowledge and evidence. My knowledge and evidence leaves open some range of multiple possible futures, and the different actions available to me will narrow down this range in different ways. The practically rational plan of action is the one that yields the narrowed range I expect will be best. (Put this in the technical language of expected utility maximization, if you like. An actor certainly ought, in her planning, to *weight* the desirability of each open future by the probability it is actual. I use ‘expect’ in its semi-technical sense to flag this.) But, for those futures that I already know are non-actual, I need not consider them in my planning; I can ignore them as I choose the action that I expect will yield the best outcome. All this is familiar and uncontroversial for rational plans of action. What’s less realized is that all this goes not only for what I plan to *do*, for what *action* I take, but for what I will *believe* as well. Just as the practically rational action yields the best expected results, it is practically rational—or, we can say *beneficial* instead of practically rational—to reason, to form beliefs, in those ways that will yield the best expected results, the desired result being actually true belief (see Joyce (1998)). Even if we have no voluntary control or guidance over our reasoning, there is still a benefit to reasoning in ways that yield the best expected outcome. It is thus beneficial to reason in ways that preserve truth throughout the range of possibilities which are not known to be non-actual. And, it is *not* similarly beneficial for my reasoning to preserve truth in those possibilities known to be non-actual.<sup>10</sup>

There are countless such possibilities that it yields no expected benefit for my reasoning to preserve truth in. We already saw a stark illustration of such possibilities when we considered interpretational validity: I don’t care that my inference that NY won still preserves truth in a metaphysically possible world where I am speaking another language or thinking different thoughts. I likewise don’t care that my inference still preserves truth in a possible world where it rained on this actually sunny day today, or a world where I’d slept all day today, or one where my parents named me Mugsy,

---

<sup>10</sup> I’m putting my claims in terms of what I *know* to be (non-)actual, rather than in terms, also commonly used in expected utility theory, of what is (*in*)consistent with my evidence. In talking this way, I don’t mean to be taking sides on any debates over the norms for practically rational action; in particular, I don’t mean to be presupposing the view of Hawthorne and Stanley (2008), who argue that there are irreducible knowledge norms, norms not reducible to any part of expected utility theory. (Moss (2013) shows one way of synthesizing expected utility theory with knowledge norms for rational action.) The reason I’m talking in terms of what I know to be (non-)actual, rather than what’s (*in*)consistent with my evidence, is to avoid having to get into the question of what notion of *consistency*, and thus *validity*, would be appropriate to use in the theory of rational action, a tricky question in the present context. Otherwise, I’m happy to use the standard framework of expected utility theory.

much less do I care about worlds where donkeys talk—all these worlds I know are non-actual. Although metaphysically valid inference insures me against failure of truth-preservation even in these possibilities, these possibilities are among the many that it's of no value for me to prepare for. Again, I know these possibilities are non-actual, and the rational (or beneficial) thing to do (or believe) depends only on what will produce favorable outcomes in the futures my knowledge and evidence leaves open—in particular, I take the action and form the belief with the best expected results. It gains me nothing to plan to produce a favorable result in possibilities already ruled non-actual by my knowledge and evidence.

Some theorists, e.g. an advocate of Jeffrey conditionalization (Jeffrey (1965/83: chapter 11)), *may* argue (Jeffrey conditionalization by no means requires this) that we can never rule out *any* metaphysical possibility—*every* contingency, and thus every metaphysically possible world, should always receive a positive chance of being actual, and thus should, to some extent, be factored into one's planning. Still, note that even on such a radical view, there is a rapidly vanishing value to the truth-preservation metaphysical validity gives you in the progressively less and less likely worlds. And, anyway, I find it highly implausible to say: I can't exclude even the wildest possibilities from my planning considerations, including ones where I'm speaking French or where donkeys talk, because I *don't know* these are non-actual, or because they are (in some interesting sense) *not inconsistent* with my evidence.<sup>11</sup>

I'm *not* claiming it's any *defect* of a metaphysically valid inference that it secures truth-preservation in the known-by-me-to-be-non-actual worlds; securing this gratuitous measure of truth-preservation isn't any cost to me. All I'm claiming here is that something good or valuable about my inference (that NY won) has been explained (in sub-section 3.2) *only* by its preserving truth in a certain *partial range* of cases. The given explanation will hold *regardless* of whether the inference preserves truth in *all* cases.

Some may be tempted here to object by saying that I have knowledge that my inference preserves truth in those worlds I'm weighing in my planning because I inferred it from knowledge of truth-preservation in all worlds, and *therefore* the fact of metaphysical validity must be a part of the present explanation. This is the mistaken objection that I addressed earlier (in sub-section 1.3). Even if we know an explanation, *E*, of a datum, *D*, by inferring it from some other fact, *F*, this does not thereby make the latter

---

<sup>11</sup> Many of Jeffrey conditionalization's fans like it because the classical alternative makes evidence certain and thus, in the standard framework, indefeasible. These fans are willing to give up the intuition that sometimes we can be rationally certain of something we've learned in order to salvage the intuition that everything we believe is defeasible. But, these same fans might now prefer the framework developed in Titelbaum (2013), a framework that allows propositions to be both certain and defeasible.

fact *F* part of the explanation of *D*. And, again, it's also not clear that you must entertain and know a premise about necessary truth, truth in all possible worlds, in order to know that the inference preserves truth in every possibility you are weighing in your planning.

It may still seem there is a natural way to explain the value of truth-preservation in these many known-by-me-to-be-non-actual possibilities, even while they are irrelevant to *my* planning, even while there's no benefit to *me* that my *present* inference (that NY won) preserves truth in these worlds *I* ruled out. It may seem the explanation begins as follows: not everyone is so lucky to know that I didn't sleep all day or that my parents didn't name me Mugsy, so *other people* may want to insure *their* reasoning against failure of truth-preservation in these cases, cases *they* do not know to be non-actual; even my past self was not so lucky as to know today's weather, so it benefited *me in the past* when my past inferences were insured against failure to preserve truth both in futures where it rains today and in futures where it doesn't rain today.

I believe this is the start of a correct explanation of a value of a certain kind of truth-preservation, but we need to understand it carefully. Even though it is desirable for certain *other* inferences to preserve truth whether or not it rains today, that doesn't mean that it's desirable for *my* inference that NY won to do so. And, the fact that *my* inference is metaphysically valid is not itself an explanation of why some *other* inference preserves truth in any given case, or of why that other inference's truth-preservation is a good or valuable thing.

I *do* want to eventually say that part of what explains the value of *my* inference that NY won is that other, *similar* reasoning preserves truth whether or not it rains today. But the sought-after explanation must mention something about a similarity relation between the two inferences. Here I'm only emphasizing that *my inference's metaphysical validity* is not what provides this explanation. To provide the explanation, we turn, in the next section, to a notion of validity where my inference's validity is partly constituted by certain other, similar truth-preserving inferences. This is the notion of substitutional validity.

## 4. Substitutional Validity

### 4.1. Characterizing Substitutional Validity

To arrive at a notion of *substitutional validity*, we understand a case as a *substitution instance* of an interpreted sentence, or of any other suitably structured representational state, such as a belief. As when we discussed interpretational validity, the mentalese hypothesis makes it easiest to assign beliefs structure, while those who think beliefs intrinsically lack the



requisite structure may still assign them structure by courtesy of structured sentences that express the beliefs in some designated language, such as the idiolect of the subject, or of the that-clauses used by an evaluator attributing validity. A substitution instance of a given interpreted sentence is any other *interpreted* sentence that uniformly replaces all or some (or none) of the simple sentences occurring within the given sentence with any other (either simple or complex) sentences. Here, as before, ‘uniformly’ means that sentences occurring twice must get the same replacements. A simple sentence contains no logical operators. The replaced sentence must be simple so as to prevent, for example, ‘Dogs bark’ from being a substitution instance of the sentence, intuitively a logical truth, ‘Dogs bark or dogs don’t bark’.<sup>12</sup>

The notion of a substitution instance extends naturally to beliefs, with the details depending on how you prefer to assign structure to beliefs. Applied to belief, the substitution instances are *other beliefs*, rather than interpreted sentences. And, the notion of substitution instance then naturally extends to inferences. An inference, which is composed of beliefs, has other inferences, inferences with different contents, as its substitution instances.

The key idea is just that, whereas interpretational validity involved varying the *contents* of some vehicle, like a sentence or a belief, substitutional validity involves varying the *vehicles* themselves, *but the vehicles always keep all their contents fixed*. The content/vehicle distinction is crucial to understanding what’s distinctive about substitutional validity.

As with interpretational validity, substitutional validity will presuppose one or another enumeration of the logical constants, and our purposes again allow us to remain agnostic among the various accounts of what, if anything, explains why a term is a logical constant.

(You might attempt to explain what makes a term a logical constant by drawing on the coming explanation of the value of an inference that’s due to its being substitutionally valid. I’ll leave such an attempt for another time; it’s not my aim in this paper to argue over logic’s boundaries. The point now is that our definition of substitutional validity doesn’t presuppose any special account of the logical constants, and is consistent with the standard approaches, including those surveyed in MacFarlane (2009).)

(Note that substitutional validity will be, more apparently than metaphysical validity was, a formal notion. Again, while formality is an important

---

<sup>12</sup> The definition given here is that of Quine (1970/86: pp.50–1). As Quine himself says, some technical details are omitted from this definition. A better, though more cumbersome, definition would talk not of sentences but well-formed formulas with certain restrictions on which of their variables can be free or bound. Details can be found in McKeon (2004: p.207). Omitting these details will be fine for our purposes, since our focus is on Modus Ponens, which obviously doesn’t generally contain quantifiers or variables.

feature for some philosophers of logic, we are not assuming any formality requirement (among the many that could be distinguished).)

The notion here of an interpreted sentence, or of a belief, can be understood as referring to a *type*, or to a *token* (token utterance or inscription, or, for belief, token mental state). I intend the interpreted sentences and beliefs that constitute substitution instances to be understood as types, where these types are individuated by the contents they carry. That is, two token interpreted sentences or beliefs are tokens of the same type of interpreted sentence or belief just in case they have the same content.<sup>13</sup>

A given sentence, belief or inference is substitutionally valid if, and only if, every type of sentence, belief or inference (respectively) that is a substitution instance of it is truth-preserving. The substitution instances must be *actually* truth-preserving, that is, truth-preserving when the contents of these (interpreted) sentences, beliefs or inferences are evaluated for truth-values in (or “at”) the actual world. That’s our definition of substitutional validity, a reasonably non-technical description that will be adequate for our purposes of evaluating whether being substitutionally valid helps explain why an inference is a good inference.

A cautionary warning: some readers may find it natural to associate the substitutional validity of an inference with the validity of a *schema* which the inference falls under. There is, however, a danger this will cause confusion or misunderstanding. Different things can be meant by talk of the validity of a schema. One approach defines validity for schemas derivatively, in terms of validity for sentences: the schema is defined as valid just in case all sentences falling under it are valid. This derivative approach to applying the notion of validity to schemas does resemble our characterization of substitutional validity. But, it will involve some separate, non-derivative characterization of validity for sentences, and this can equally well be given however you like, whether interpretationally, metaphysically, substitutionally, or however else. So, since my aim is to advocate an understanding of validity that is thoroughly substitutional, it doesn’t much help things to bring in schemas where validity is derivatively defined for them. An alternative approach to defining validity for schemas is to give a non-derivative definition, but if we go this way, we’ll find it’s just as easy to give the definition either in an interpretational way or in a substitutional way—do the schematic letters get reinterpreted, or are they substituted by letters with fixed interpretations? For an illustration, see Quine (1970/86:

---

<sup>13</sup> Technical detail: don’t individuate in a way that includes *de se* contents so fine-grained that some self-ascriptions have contents that no other token belief can share. Throughout, I’ve avoided discussing context-sensitive aspects of content, assuming that context-sensitivity raises no insoluble problems for my purposes in this paper. However, this point matters to what I say below. See footnote 17.

pp.49–52), where he non-derivatively defines both model-theoretic (interpretational) and substitutional validity for schemas. So, it isn't likely to be much help, and it is likely to cause misunderstanding, to understand the substitutional approach to characterizing validity as being just a schematic approach to characterizing validity.

Let me, without further delay, indicate that I'm aware that some famous objections have been raised in the past against substitutional accounts of validity on technical grounds. Indeed, Tarski (1936/1983) himself argued in favor of his model-theoretic (interpretational) account of validity primarily because, he argued, a substitutional account is inadequate. Substitutional accounts of validity were in the air prior to Tarski's model-theoretic account. Tarski criticized Carnap for characterizing validity substitutionally.<sup>14</sup> Bolzano is famously given credit for a very early substitutional account of validity.<sup>15</sup> In fact, it is not so widely recognized, but *Frege* briefly sketched and endorsed the notion we're here calling substitutional validity, though he did not seem to endorse it as an analysis of validity, but rather as a necessary and sufficient condition of validity, and thus as an appropriate method of testing for validity.<sup>16</sup> (He favored substitutional validity, which takes all terms to have fixed interpretations, because, recall, he took issue with Hilbert's practice of re-interpreting terms.)

I believe the alleged technical problems with substitutional validity can all be satisfactorily addressed. Quine was a major advocate of understanding validity as substitutional validity (though for reasons completely different from the ones to be given here) and, he has, along with several commentators, elaborated responses to the technical problems. I'll delay and confine discussion of, and responses to, the technical problems to sub-section 4.4. I think we don't need to get into technical details, however, to see the intuitive argument for why substitutional validity best explains what is good or valuable about a valid inference. So, let's turn to that right now, and hear the real answer to our main question.

---

<sup>14</sup> Friedman (1988/99: p.171) usefully distills the definition of Carnap (1934/2000): 'Carnap proceeds on the basis of a prior distinction between *logical* and *descriptive* expressions (§50). . . . Given the distinction . . . , we then define the analytic (L-true) sentences of a language as those theorems (L- or P-consequences of the null set) that remain theorems under all possible substitutions of *descriptive* expressions (§51).'

<sup>15</sup> Etchemendy (1990) presents Tarski's notion as the natural improvement upon one that he attributes to Bolzano. Quine (1954/76: p.110) also gives credit, by hearsay, to Bolzano for having a substitutional account of validity. I find it doubtful how similar Bolzano's own view is to that presented here; see Sher (1996: p.662 including note 14) and Rusnock and Burke (2010: especially p.19) on this, but the historical issue isn't important here.

<sup>16</sup> For brief exposition and attribution of the view to Frege, see Resnik (1974: pp.398–9) and Tappenden (2000: p.274), interpreting brief remarks in Frege (1906/84: section III). Resnik explicitly notes that Frege's view is like Bolzano's and Quine's, Frege apparently having arrived at it independently.

#### 4.2. Substitutional Validity Explains Why a Valid Inference Is Good

Unlike with interpretational validity, substitutional validity isn't defined in terms of what a given inference, or a given reasoner, *could have meant or could have been thinking about*, though we know she actually does not. And, unlike with metaphysical validity, it isn't defined in terms of the inference's preserving truth in *other worlds*, many of which we know are non-actual. The substitutional validity of a given inference is defined in terms of other content carrying vehicles that we might or others might deploy. It is defined in terms of other inferences that share the form or structure of the given inference, inferences that actual people might actually perform, and that preserve truth in the actual world. Substitutional validity gives us what we've been looking for, and what we did not find in metaphysical validity at the end of our discussion in section 3: it offers a way of linking *my present* inference's truth-preserving properties to the truth-preserving properties of *other* inferences. The link is right there in the definition of a substitution instance, and thus of substitutional validity. The substitutional validity of my inference that NY won *just consists* in the truth-preservation of all *similar* inferences, in a specific sense of 'similar' that captures just the inferences by Modus Ponens. This close definitional link to other inferences, inferences that we actually do and should care about, is what allows substitutional validity to be a variety of validity that can answer our original question.

Our original question was: we know what's good about an inference's being truth-preserving, but beyond that how does an inference's being valid contribute to its being good? Here is my answer. I suggest that substitutional validity is a good feature of an inference because, when an inference is substitutionally valid, it exemplifies a *way of reasoning* that is worth *recommending*. A valid inference is a good inference because it is an instance of a commendable way of reasoning. A 'way of reasoning', as I mean it here, is all reasoning that shares a logical form; in our running example it's Modus Ponens. So, the sense of 'good' or 'valuable' in which a substitutionally valid inference is good or valuable is that the inference is commendable or *praiseworthy*, and that's clearly a way of being good or valuable since it's (quite literally) a way of being worthy.

The kind of value that I'm suggesting is explained by an inference's validity is different from the kind of value that's due to a belief's being true or an inference's being truth-preserving. By virtue of being true, my true beliefs tend to *cause* the satisfaction of my desires. But, an inference's being substitutionally valid is not itself a cause of, or even evidence of, its substitution instances' being truth-preserving. So, unlike the value of truth and truth-preservation, the value of an inference's validity is not a similarly instrumental value. But, this does not mean that substitutional validity is not a good or worthy thing. It is a praiseworthy thing.

Is there still the problem we saw with metaphysical validity, namely that we should not really care about *all* the other inferences being truth-preserving? No, there is no problem this time. Clearly, we should care about the many alternative cases that we know are instantiated by *actual* inferences. When I actually infer, say, that Boston lost on the basis of premises that Boston lost if Chicago won, and Chicago won, my reasoning and my beliefs actually instantiate an alternative case, one that partly constitutes the substitutional validity of the original inference about NY and LA. Of course, there will be many cases constitutively involved in the validity which are merely possible inferences; they're never actually performed, never actually instantiated. However, even though we know there exist cases that are never actually instantiated, it is still important to us, it is still valuable, that all cases be truth-preserving. Why? Why should we care whether the never-actually-instantiated cases are truth-preserving? Because, when it comes to substitutional validity, there are no cases, no substitution instances, that we *know* will never actually be performed.<sup>17</sup> Although I know the possible world where I'm speaking French is non-actual, you and I know nothing that can restrict what kind of reasoning other people, or other reasoners of whatever kind, will engage in down the line. We cannot even put an upper bound on how complex the reasoning might get, especially given the ever surprising pace at which computers can perform increasingly complex tasks, some of which, I'm willing to claim, clearly count as inferences, at least for our evaluative intents and purposes. This is in contrast with interpretational validity, where we *know* of all the alternative re-interpretations of our inferences that they are non-actual; and it is in contrast with metaphysical validity, where we *know* of many (though not all) of the possible worlds that they are non-actual. It is thus good that an inference preserve truth in *all* substitution instances.<sup>18</sup>

---

<sup>17</sup> Technical worry: don't we know that there is no actual performance of Modus Ponens that involves the *de se* content that Abraham Lincoln's thought would have had if he had self-ascribed the property of living past age sixty? Reply: as mentioned in footnote 13, I treat this as an inference of the same type as an inference that might be performed by someone else, such as someone who lives past sixty.

<sup>18</sup> I should address a technical issue, though: the semantic paradoxes, such as the Liar and Curry's Paradoxes, create a problem for the claim that *all* substitution instances of Modus Ponens preserve truth. Still, any plausible solution to the semantic paradoxes will say that *practically* all instances of Modus Ponens preserve truth. We could still have raised our question, and reached our answer, in this paper more or less as we did. The question becomes: why should we value not only a Modus Ponens inference's truth-preservation, but its validity, its truth-preservation in *so many other cases* (even if not *all* other cases, given the paradoxes)? Our answer remains: because we care about truth-preservation in other inferences we, or others, may perform. See Field (2006, 2009b) for presentation of the technical problem, and discussion sympathetic to the attitude I'm recommending.

### 4.3. *The Value of Validity, and the Value of Our Evaluative Practice*

That is my answer to the question of the paper: how does a valid inference's validity in particular explain why the inference is a good inference? And (as I promised in section 1) the answer was not given in terms of epistemic rationality. Now, however, in this sub-section, I want to make an additional but complementary point, one that does concern rationality, or rather the concept thereof. For, my proposed explanation of the value of a valid inference was that it has value in the form of praiseworthiness, and the usual way that we in fact praise an inference that is worthy of praise is by calling it rational. So, now, what I want to show in this sub-section is that my proposal about the value of an inference's validity naturally fits together with a plausible general view of *why* we make epistemic evaluations in the way we do. It's a view that I think is supportive of the explanation I've proposed for the value of validity, but the two are, strictly speaking, independent (it conceivably could be that just one is true), so I will just very briefly sketch this plausible general view of epistemic evaluation and I'll situate my proposed explanation of the value of validity within it. (This view is from Dogramaci (2012, forthcoming), where I named it *epistemic communism*.)

We evaluate reasoning as epistemically rational or irrational. We do this using not only 'rational' but any number of ordinary terms capable of expressing this same meaning, words like 'good', 'clever', 'logical', 'bad', 'stupid'. And, this practice can be given a plausible teleological explanation, that is, an explanation of why we engage in the practice stated in terms of an interest of ours that it promotes. We engage in this practice *because* we have an interest in promoting and suppressing, in ourselves and especially in others, certain future inferences. For, even though the praiseworthy substitutionally valid inference normally does not *by itself* cause its substitution instances to be performed, our positive evaluation of it, our praise of it as rational, can cause, via our encouragement, more of these inferences of the same truth-preserving sort. We want to see more of the reasoning we praise, and less of what we criticize, and what we specifically want in that reasoning is truth, *actual* truth. And we want to see this not only in our own future reasoning, but in the reasoning of others, since we rely on testimony for so much of what we know. (I even think 'rely' and 'so much' grossly understate the situation: the reliability of *almost all* of what you believe is, even if not directly then indirectly, at the *mercy* of others' reliability. Even if you could somehow survive without reliable testimony, it's unimaginable to me how you could live a good life.)

Now, when we make epistemic evaluations, when we engage in this activity that I've just emphasized can have such important causal effects, the *object* of our evaluation is, according to many epistemologists, a

*belief-forming rule*.<sup>19</sup> And it's here where we can see how epistemic evaluation and substitutional validity are so closely connected, for, substitutional validity just is truth-preservation throughout the range of inferences that all fall under a common belief-forming rule. In our main example, the substitutional validity of the single inference that NY won just is the actual truth-preservation of every other inference by Modus Ponens, any of which might be or become actually instantiated. So, if the way we epistemically evaluate an inference is according to the rule it employs, and we thus take an interest in what rules subjects are employing in their reasoning, that would then very naturally explain our interest in this notion of validity, substitutional validity, which concerns all the inferences—any of which might be actually performed—that fall under a common rule. When a subject does well, or poorly, with a token inference, we comment on the success, or failure, *of the rule* because this is what bears on future occasions, actual future token inferences, inferences which we hope will do well with regard to the thing we value, truth. And substitutionally valid inference just is an inference that does well with regard to truth: it does well both itself and in the other inferences that are other instances of the rule that it falls under.<sup>20</sup>

But, be careful not to confuse together two distinct, though complementary, things. The explanation of *why we positively evaluate an inference for being valid* (just sketched in this sub-section) is that such a positive evaluation has good effects; going around evaluating in this way is an instrumentally valuable thing to do. The explanation of *why the inference is valuable for being valid* (given in the last sub-section before this one) is that validity, substitutional validity, is praiseworthy. A substitutionally valid inference is a good inference not because its substitutional validity is instrumentally valuable, but rather because its substitutional validity is praiseworthy. The substitutionally valid inference is *worth* positively evaluating because—as the definition of substitutional validity tells us—the inference exemplifies a form, or a rule, every instance or application of which is a truth-preserving inference.

---

<sup>19</sup> Epistemologists who argue that the objects of epistemic evaluation are belief-forming rules (methods, processes, policies—use whichever term you like), include Pollock and Cruz (1986/99), Goldman (1986), Wedgwood (2002), Peacocke (2004), Boghossian (2008) and Field (2009a).

<sup>20</sup> Views similar to this general view of our evaluative practice can be found in some literature, including Field (2009a), and, in a more general context than just epistemic normativity, Gibbard (1990, 2003). These works are noted for the *non-cognitive* character of the views they endorse, but I intend to be neutral with respect to that particular aspect of the views. I also don't intend to take sides with anti-realism. (Gibbard (2003: especially the preface) indicates why his view is not clearly anti-realist.) Rather, I mean to be favoring the way in which these views theorize about the *function*, in particular the *advice-giving* function, of normative evaluation. See especially Field (2009a: p.286) and Gibbard (1990: pp.72–80, 223–6).

I can now briefly return to a point I raised early on: as Harman observes, only a select number of valid inferences receive our praise as rational, while most valid inferences, those that leap to remote consequences, are criticized as irrational. What explains this discriminatory practice? My view here has been that *all* substitutionally valid inferences are praiseworthy. They all exhibit the kind of reliability that is worth caring about and promoting. However, due to practical limitations, creatures like us are built such that we can immediately recognize, without a proof, the validity of only a few basic rules. For us, these basic rules are Modus Ponens and a handful of others. It requires, for us, a tremendous amount of effort to produce the proofs that allow us to rationally recognize the validity of most other valid rules; consider, for example, a rule for inferring Fermat's Last Theorem directly from the Peano Axioms. My larger view is that, while all valid rules are equally praiseworthy for being valid, practical limitations forced us to adopt just a small handful of valid basic rules, and it is *arbitrary* which few got adopted and are thus praised by us. Nothing explains why, among all the praiseworthy rules, only Modus Ponens and a handful of others earn our praise as rational. It would have been just as well if we were built to use and were to praise some other set of praiseworthy valid rules (though we'd need some other term of praise, since, I believe, 'rational' rigidly picks out its extension). For elaboration and defense of this claim of arbitrariness, see Dogramaci (forthcoming). I argue there that it is inexplicable, because it is in a sense conventional, that Modus Ponens is rational while a leap to Fermat's Theorem is irrational, even though both are valid and thus, according to my view here, equally praiseworthy.

#### 4.4. *Technical Issues with Substitutional (and Interpretational) Validity*

In this sub-section, I review two well-known lines of objection to substitutional validity on technical grounds, and I describe the (not sufficiently well-known) correct responses. The first line of objection concerns sentences whose logical truth seem to be sensitive, in a counter-intuitive way, to the availability of substitute terms elsewhere in the language. The second line of objection concerns sentences that assert the existence of specific numbers of objects.

In this section, I drop my focus on Modus Ponens. Doing this is required in order to state the second objection, and it also makes the first objection easier to discuss. Although this paper's main aim has been to argue for an answer to a question focused on Modus Ponens, the real interest of the issue clearly depends on being able to generalize the argument more widely, and thus the issues raised and addressed in this section are important to my broader purposes.



(1) *Sensitivity to the Lexicon of a Language.* One line of criticism, voiced by Tarski (1936/1983: pp.414–6) (against Carnap (1934/2000)) and by Hinman, Kim and Stich (1968) (against Quine (1936/76, 1954/76)), is that a substitutional approach will count an intuitively invalid sentence as valid *if there is a sufficient shortage in the stock of terms that we may substitute into the sentence.* A simple example from McKeon (2004): we'll end up counting as valid the sentence ' $\forall x\forall y(Fx\leftrightarrow Fy)$ ' if all the predicates in our language, ' $F$ ' and all the rest, apply to everything; a substitution instance falsifying that sentence requires a predicate that applies to just some things.

The real objection here is not simply that the substitutional definition of validity risks misclassifying a sentence as a logical truth. The deeper concern is that the substitutional approach makes a sentence's logical status sensitive to what terms show up, not in the sentence itself, but in the rest of its language's lexicon. Intuitively, the sentence's logical status should be fully settled by the terms in the sentence itself, independently of the language we take it to be a part of.

This objection is avoided if we characterize substitutional validity according to Quine's considered view. In Quine (1970/86: p.59), the final proposal is stated as follows: '*a logical truth is a sentence that cannot be turned false by substituting for lexicon, even under supplementation of lexical resources.*' The idea here is that the test for a sentence's validity is not sensitive to what terms are in any one language, because, on this proposal now, even the terms that are *not* in that language are equally relevant to whether the sentence is valid.

Quine wanted to think of our language's lexicon as having finitely many terms, though there are infinitely many ways of extending a lexicon. A subtly distinct proposal would define validity relative to *one* specific language whose lexicon contains *all* terms. Quine had his own idiosyncratic reasons for wanting to think of our language's lexicon as finite, which we'll see shortly below.

McKeon (2004: p.321, note 17) points out that, by making the lexicon indefinitely extensible, we also neutralize an objection raised by Boolos (1975/98: pp.51–3) against substitutional validity. Boolos constructs a set of sentences that is satisfiable (model-theoretically), but he shows it has no true substitution instances. However, Boolos's construction assumes the substitutions must come from a fixed lexicon.

(2) *Asserting the Existence of a Number of Things.* The second well-known criticism of substitutional validity is that there seem to be cases where substitutional validity diverges from interpretational, specifically model-theoretic, validity for first-order logic, and substitutional validity appears to be giving the wrong verdicts. The suggestion that there is divergence is a bit interesting and even surprising, because Quine (1970/86: chapter 4) claims to *prove* there will be no divergence, as long as the

substitutionalist may presuppose an ontology, i.e. a range of objects that the quantifiers quantify over and the terms can refer to, that contains at least the natural numbers (and as long as our lexicon contains the language of arithmetic, but we already agreed to let the lexicon be indefinitely extensible). So, if Quine alleges such a result, how could there be any worry about divergence?

The *apparent* cases of divergence involve assertions of the existence of various numbers of things. Take for example, ‘There are two things’, that is, ‘ $\exists x\exists y\neg x=y$ ’. It’s clearly not a model-theoretic logical truth, because the sizes of the domains of models vary, and some have size one. But, on the standard characterization, including Quine’s, substitutional validity makes no provision for a varying domain, so this would *appear* to be a substitutional logical truth. (Sher (1996: 663–4) and Hanson (1997: section 1) cite this as a problem for substitutional validity).

Quine is aware of the situation. In chapter 5 of Quine (1970/86), he observes that the divergence arises only if the identity predicate is treated as a logical constant, a predicate not replaced in the sentence’s substitution instances.

The simplest way for the substitutionalist to match her classification of validity to the model-theoretician’s is to declare that both the substitutionalist and the model-theoretician should treat the identity predicate, like any other two-place predicate, as a non-logical term. However, this strategy is not ideal; it limits the power of our logical theory, because now we cannot say that ‘ $a=a$ ’ is a logical truth, and, if possible, we’d like to be able to mimic the model-theoretic treatment of that as a logical truth (as happens when the model-theoretic view treats the identity predicate as a logical constant).

Quine wants to treat ‘ $a=a$ ’ as a logical truth, and so he gives a different, more clever solution to the problem. Quine proposes that ‘=’ is a non-logical term, but it is not a simple two-place predicate. Rather, any formula of the form  $x=y$  is an abbreviation for a formula that asserts, in effect, that no predicate of the lexicon discriminates between  $x$  and  $y$ . How this is achieved depends on what predicates are in the lexicon. If there are only one-place predicates, we need only replace  $x=y$  with a conjunction of biconditionals of the form  $Fx\leftrightarrow Fy$ , one such biconditional for each predicate in the lexicon. If there are polyadic predicates, it just requires using a few universal quantifiers, e.g.  $\forall z[(Rzx\leftrightarrow Rzy)\&(Rxz\leftrightarrow Ryz)\& \dots]$ . This, by the way, is the reason mentioned above that Quine wants a finite lexicon.

Quine’s technique succeeds in classifying examples like ‘ $\exists x\exists y\neg x=y$ ’ as not logical truths, for now they have false substitution instances (just replace all the predicates that ‘=’ is abbreviating with, say, empty predicates). At the same time, this technique succeeds in making substitutional logical truths out of the usual axioms for identity, namely all instances of

self-identity and all instances of Leibniz's law. It thereby mimics exactly the standard model-theoretic classification of logical truths, and valid arguments, for first-order logic with identity as a logical constant.

A shortcoming of Quine's solution is that, although it now correctly classifies the logical truths, it does not appear to assign '=' its *intended* extension. As intended, there might be two distinct things which no predicate discriminates. Even though Quine wasn't bothered (he wanted ontology and language to match more closely than we might be happy with), we might want a better solution.

A solution is proposed by Berlinski and Gallin (1969) and McKeon (2004). They describe the same basic idea. I'll follow McKeon's presentation. The suggestion is that we leave '=' as a logical term with its fixed, standard interpretation, and instead we make a different, slight modification to our definition of substitutional validity, but one that remains in the substitutionalist and Quinean spirit. The substitutionalist can, in a natural way, achieve just what the model-theoretician achieves by varying the domain. She can do this by treating a sentence's variables as non-logical terms that get replaced across substitution instances, and having the variables in each of the different substitution instances quantify over domains of different sizes (each individual sentence's variables sharing the same domain, of course). (Alternatively, an equivalent way of implementing this solution—just a notational variant—is to treat the quantifier symbols as non-logical terms.<sup>21</sup>) To allow these new replacements, liberalize the earlier restriction that a sentence's substitution instances only replace its simple sentences; quantified sentences may be replaced, but only by other quantified sentences that differ just in the replacement of the variables.

To illustrate this solution, consider ' $\exists x\exists y\neg x=y$ ' one more time. McKeon usefully suggests we subscript each sentence's variables with the size (cardinality) of their domain. Then, for example, ' $\exists x_1\exists y_1\neg x_1=y_1$ ' is false, but all its substitution instances, such as ' $\exists x_2\exists y_2\neg x_2=y_2$ ' and ' $\exists x_{N_0}\exists y_{N_0}\neg x_{N_0}=y_{N_0}$ ', are true. (This proposal is in the Quinean spirit, because Quine's concern was to minimize the set theoretic commitments of our theories, and this proposal need not rely on any more sets than Quine's own proposal did; see McKeon (2004: p.220).)

This suffices to show that substitutional validity can match whatever verdicts interpretational validity can give, at least for first-order logical truth. However, this may seem to now raise a philosophical concern. If we can convince ourselves that substitutional validity and interpretational validity are co-extensive, then how can one, but not the other, explain why a valid

---

<sup>21</sup> Shapiro (1998: p.143, note 12) observes that Enderton (1972: pp.70, 81) counts the universal quantifier as a non-logical term which different models assign different extensions (namely, domains).

inference is a good or valuable inference? Was it rash to conclude, in section 2, that interpretational validity does not explain what's good about a valid inference? I address this issue in the next, and final, sub-section of the paper.

#### 4.5. *Why Substitutional Validity Provides the Best Explanation*

It should already be clear that there is a potential gap in the philosophical significance of co-extensive notions: Quine demonstrates this when he explicitly argues both (a) that substitutional and interpretational validity are co-extensive, and (b) that we should ultimately prefer a theory that speaks of substitutional validity (described as such), but not interpretational validity. But, Quine's reasons for this concerned ontology, while we are concerned with the normative character of validity. So, in this sub-section, let's suppose substitutional and interpretational validity indeed are co-extensive, in fact necessarily so, and I'll argue that two necessarily co-extensive notions do not generally have the same explanatory significance, and that our pair in particular do not.

(Note, I'm not presupposing, here or anywhere in the paper, any particular general theory of explanation. I have no theory of explanation in mind. My main aim in this paper is only to engage in the first-order practice of giving explanations, which, I trust, it's fair to do without a theory of explanation in hand.<sup>22</sup>)

Our two co-extensive notions, substitutional and interpretational validity, have different definitions, and are thus different notions (or concepts, or understandings).<sup>23</sup> This is what allows them to differ in what they explain. Explanation is intensional. Now, it's true that this claim won't be convincingly illustrated by just any pair of necessarily co-extensive notions. It might seem that some explanations that use 'equilateral triangle' (or the concept those words express) would be just as good if we replaced that with 'equiangular triangle' (or the associated concepts). But, a well-chosen example will better support our claim that explanation is intensional. Take a pair

---

<sup>22</sup> If you really want, maybe for illustration, a concrete example of a theory of explanation that straightforwardly delivers the claim that necessarily co-extensive notions do not generally have the same explanatory significance, one such theory would be that of van Fraassen (1980: chapter 5), according to which a candidate explanation's power or success is sensitive to the interests of a (non-ideal, non-omniscient) inquirer. On p.153 van Fraassen even explicitly favors the suggestion that truth-conditionally equivalent propositions can have different relevance properties and hence different explanatory properties.

<sup>23</sup> Don't get tripped up on the fact that, for both interpretational and substitutional validity, cases are types that are *individuated* by sameness of content. Individuating by content is just a way of sorting. You can sort both marbles and playing cards by sameness of color, but marbles aren't cards. Likewise, the cases defining interpretational validity aren't the cases defining substitutional validity.

of expressions (or concepts) that are necessarily co-extensive, but the co-extension is highly non-obvious. Let's pick another mathematical example.<sup>24</sup> Take, for example, the surprising equivalence that turns out to exist between the properties attributed to sets by the Axiom of Choice and the Well-Ordering Theorem. The Well-Ordering Theorem states that, given any set  $S$ , there exists a well-ordering of  $S$  (a linear ordering where every non-empty subset of  $S$  has a least element). Suppose I wanted to explain why a certain particular set, say, the real numbers (set-theoretically characterized), can be well-ordered. One explanation is simple: this is just an instance of the generalization stated by the Well-Ordering Theorem. The following alternative explanation is *not* equally good: given any set  $S$ , there exists a choice function for  $S$ , that is, a function that returns a member for each non-empty member of  $S$ . That's just what's stated by the Axiom of Choice. It turns out to be equivalent to the Well-Ordering Theorem; the property of having a "choice function" is equivalent to the property of being "well-orderable". But obviously, if you just state the Axiom of Choice here, you will not provide an equally good explanation of what was explained by the Well-Ordering Theorem. More generally, it's not true that any necessary and/or sufficient condition for some explanation can itself provide that same explanation.

What if it's *known* that something is a sufficient or even equivalent condition? Can those who are *in the know* give their own explanation using the sufficient or equivalent condition? This doesn't seem like it will generally work. If I complicate my explanation that just mentioned the Well-Ordering Theorem by adding two more claims, (i) the Axiom of Choice, and (ii) the claim that the Axiom of Choice entails the Well-Ordering Theorem, I'll sacrifice simplicity at no apparent gain in my explanation. Maybe the result of this can still be called an explanation in some sense, but it's clearly not the best explanation.

If we were comparing two equally simple explanations, maybe one stated using Well-Ordering but not using Choice, and another using Choice but not Well-Ordering, then perhaps there would be a debatable issue. But, we don't generally have this option. Again, just mentioning Choice alone doesn't do the explanatory work in our earlier example.

Now compare, if we only mentioned interpretational validity, could we put together an explanation that's just as good as the explanation we provided of the value of valid reasoning in terms of substitutional validity? We already argued that alternative re-interpretations of my beliefs do not seem to help explain the value of a valid inference. But, then, what new option is there for explaining the value of a valid inference in terms of interpretational validity, except by somehow including the extra claim that an interpretationally

---

<sup>24</sup> See Mancosu (2011) for general discussion of explanation within mathematics.

valid inference is a substitutionally valid inference? You could try adding a claim that doesn't use the terms 'substitutional validity' but instead only says that an interpretationally valid inference will be such that other inferences similar to it in such-and-such ways are truth-preserving, but that looks like it would just use the definiens of substitutional validity. And, of course, mentioning interpretational validity together with any detour through substitutional validity or its definiens will not give us the simplest explanation we are already able to give. For, our explanation, stated using the notion of substitutional validity, does not make any mention of interpretational validity or the notions involved in the definition of interpretational validity.

A fan of interpretational validity might try to defend the comparative explanatory power of her notion by arguing that a *re-interpretation* and a *substitution instance* should really be thought of as one and the same thing, the same *case* under two different descriptions or conceptualizations. This is an attempt at a kind of reconciliation, one that claims that the cases of interpretational validity and the cases of substitutional validity are the very same thing. But, this is not a tenable claim. What's true is that we can *associate*, even in a one-to-one way, the possible re-interpretations of my inference that NY won and the other inferences that constitute the substitution instance of my inference. So, for example, we can associate (a) the Modus Ponens inference that no one's home because the lights are off and if they're off then no one's home (a substitution instance of my NY inference) with (b) the re-interpretation of my NY inference where I am thinking the relevant thoughts about the lights and whether anyone is home rather than thinking, as I actually did, about NY and LA's game. One can serve as a stand-in for the other. But, to *identify* (a) with (b) is simply a mistake. Substitution instances are *other inferences*, composed of *other beliefs* (or sentences), whereas re-interpretations concern *re-interpretations of the same beliefs* (or sentences), such as my beliefs about NY and LA. The re-interpreted vehicle (sentence or belief) 'NY won' is not the same thing as 'No one is home' with its actual interpretation.

This is no idle, merely verbal distinction. To appreciate why it's important to not collapse together re-interpretations of an inference and other inferences, it's important to see their different relations to the thing that we really care about, the good stuff. We care about truth, and (derivatively) truth-preservation, in our own beliefs and inferences, and (derivatively) in the beliefs, inferences and also the *testimony* of others. It was in connection to the pre-theoretically clear value of those things that we, in sub-section 4.2, showed how an inference's substitutional validity explains why it is a good inference, and we showed this so easily because substitution instances *just are* beliefs that others may have, or sentences others may use to express those beliefs in testimonial assertions—that's what substitution instances were *defined* to be.

You can, if you want, propose to treat re-interpretations of my NY inference as *stand-ins*, entities that function as indices of a sort, for those beloved other beliefs, inferences and testimonial assertions that are defined to constitute substitution instances. But still, reinterpretations themselves are not what I care about. I care about the *beliefs* and the *assertions* of someone who is, say, thinking about or talking about the lights being off and no one being home. It's what interpretations would be standing in for that I care about.

It's true that, the fan of interpretational validity is, in a way, ever so close to being able to say the same things that I want to use the notion of substitutional validity to say. All she needs to do is ask us to understand her talk of re-interpretations as functioning as talk of other inferences, to treat interpretations as stand-ins for substitution instances. Once she tells us to understand her talk in that way, once she makes the right prefatory speech, she is ready to say everything I said about why a valid inference is a good inference. But, of course, then she ends up just piggybacking on the explanation that we can give *more simply* just in terms of substitutional validity.

So then, explanations that mention interpretational validity must end up including a detour via claims about substitutional validity (or its definiens), but an explanation in terms of substitutional validity need not mention interpretational validity. The situation is asymmetric in favor of substitutional validity. Substitutional validity thus seems able to provide the simplest, and thus the best, explanation of what makes a valid inference a good inference.

## 5. A Pedagogical Conclusion

In closing, let me suggest that we reconsider the standard story now told in most introductory logic classes and textbooks. Logic, as it's taught (i.e. the study of interpretational validity), doesn't just fail to be *especially* relevant to good reasoning, as Harman has long said. Logic, as it's taught, just isn't very directly relevant to good reasoning. If interpretational validity is relevant to good reasoning, it is only by piggybacking on substitutional validity.

We may appear to have two options. Option one is that we teach substitutional validity, rather than interpretational validity. This will obviously require much effort.<sup>25</sup> It may appear that another option is to continue to

---

<sup>25</sup> Quine's own textbook, *Methods of Logic* (Quine (1950/82)), contains a presentation of the standard model-theoretic notion of validity (sections 18 & 27), but he also emphasizes from the start (p.4), and throughout the book, that the fundamental logical notions, including logical truth, should ultimately be characterized in substitutional terms. Hanson (1997: 370–1) criticizes Quine's inconsistency here. McKeon (2004: 216–7) suggests that the most charitable overall reading of Quine's views requires that we look past some of these discrepancies in *Methods of Logic*.

teach things in terms of interpretational validity, never mention substitutional validity, but instead re-write our textbooks' opening pages, and our opening lectures, so that we no longer characterize logic as most directly concerning good reasoning. But, this will raise a difficult question about why we should care about interpretational validity. Option one may be unavoidable.<sup>26</sup>

### References

- Barwise, Jon and John Etchemendy (1999). *Language, Proof and Logic*. New York: CSLI Publications.
- Beall, JC and Greg Restall (2006). *Logical Pluralism*. Oxford: Oxford University Press.
- Bergmann, Merrie, James Moor and Jack Nelson (1980/2008). *The Logic Book*. 5th edition. New York: McGraw-Hill.
- Berlinski, David and Daniel Gallin (1969). "Quine's Definition of Logical Truth." *Noûs* 3(2), 111–128.
- Blanchette, Patricia (1996). "Frege and Hilbert on Consistency." *The Journal of Philosophy* 93(7): 317–336.
- (2012). "The Frege-Hilbert Controversy." In Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy, fall 2012 edition*. <http://plato.stanford.edu/archives/fall2012/entries/frege-hilbert/>.
- Boghossian, Paul (2003). "Blind Reasoning." *Proceedings of the Aristotelian Society, Supplementary Volume* 77(1): 225–248.
- (2008). "Epistemic Rules." *The Journal of Philosophy* 105(9): 472–500.
- and Gideon Rosen (ms/2004). "Logic and Reasoning." presented at the Central Division APA 2004.
- Bonevac, Daniel (1987/2003). *Deduction*. 2nd edition. Oxford: Blackwell.
- Boolos, George (1975/98). "On Second-Order Logic." In Richard Jeffrey (ed.), *Logic, Logic, and Logic*, 37–53. Cambridge, MA: Harvard University Press.
- Burge, Tyler (1998). "Frege on Knowing the Foundation." *Mind* 107(426): 305–347.
- Carnap, Rudolf (1934/2000). *The Logical Syntax of Language*. London: Routledge.

---

<sup>26</sup> For helpful comments on this paper, I'd like to thank two anonymous referees, Paul Boghossian, Gary Ebbs, Hartry Field, Dan Greco, John Morrison, Jim Pryor, Josh Schechter (whose comments deserve a future paper to be fully addressed), Miriam Schoenfeld, David Sosa, Jason Stanley, an audience at a workshop at the London Institute of Philosophy organized by Julien Murzi and Florian Steinberger, an audience at *TEX - the Texas Epistemology Extravaganza*, an audience at Columbia University, and an audience at the Midsummer Philosophy Workshop in Edinburgh, where Kate Nolfi and Magdalena Balcerak Jackson delivered commentaries.



- Carroll, Lewis (1895). "What the Tortoise Said to Achilles." *Mind* 4(14): 278–280.
- Chalmers, David (2012). *Constructing the World*. Oxford: Oxford University Press.
- Craig, Edward (1990). *Knowledge and the State of Nature*. Oxford: Oxford University Press.
- Dogramaci, Sinan (2010). "Knowledge of Validity." *Noûs* 44(3): 403–432.
- (2012). "Reverse Engineering Epistemic Evaluations." *Philosophy and Phenomenological Research* 84(3): 513–530.
- (2013). "Intuitions for Inferences." *Philosophical Studies* 165(2): 371–99.
- (forthcoming). "Communist Conventions for Deductive Reasoning." *Noûs*.
- Enderton, Herbert (1972). *A Mathematical Introduction to Logic*. San Diego: Academic Press.
- Etchemendy, John (1988). "Tarski on Truth and Logical Consequence." *The Journal of Symbolic Logic* 53(1): 51–79.
- (1990). *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.
- (2008). "Reflections on Consequence." In Douglas Patterson (ed.), *New Essay on Tarski and Philosophy*, 263–299. Oxford: Oxford University Press.
- Field, Hartry (2006). "Truth and the Unprovability of Consistency." *Mind* 115(459): 567–606.
- (2009a). "Epistemology without Metaphysics." *Philosophical Studies* 143: 249–290.
- (2009b). "What Is the Normative Role of Logic?" *Proceedings of the Aristotelian Society, Supplementary Volume* 83(1): 251–268.
- Fine, Kit (1994). "Essence and Modality." *Philosophical Perspectives* 8: 1–16.
- Frege, Gottlob (1893/1997). "The Basic Laws of Arithmetic: Volume 1." In Michael Beaney (ed.), *The Frege Reader*. Oxford: Blackwell.
- (1906/84). "The Foundations of Geometry: Second Series." In Brian McGuinness(ed.), *Collected Papers on Mathematics, Logic, and Philosophy*, 293–340. Oxford: Blackwell.
- (1918/1997). "Thought." In Michael Beaney (ed.), *The Frege Reader*, 325–345. Oxford: Blackwell.
- Friedman, Michael (1988/99). "Analytic Truth in Carnap's Logical Syntax of Language." In Michael Friedman (ed.), *Reconsidering Logical Positivism*, chapter 7. Cambridge: Cambridge University Press.
- Gibbard, Allan (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

- (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Goldfarb, Warren (2003). *Deductive Logic*. Indianapolis: Hackett.
- Goldman, Alvin (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- (1992). “Epistemic Folkways and Scientific Epistemology.” In *Liasons: Philosophy Meets the Cognitive and Social Sciences*, 154–175. Cambridge, MA: MIT Press.
- Hanson, William (1997). “The Concept of Logical Consequence.” *The Philosophical Review* 106(3): 365–409.
- Harman, Gilbert (1986). *Change in View*. Cambridge, MA: MIT Press.
- (2009). “Field on the Normative Role of Logic.” *Proceedings of the Aristotelian Society* 109(1): 333–335.
- Hawthorne, John and Jason Stanley (2008). “Knowledge and Action.” *The Journal of Philosophy* 105(10): 571–590.
- Heck, Richard (2007). “Frege and Semantics.” *Grazer Philosophische Studien* 75: 27–63.
- Hinman, Peter, Jaegwon Kim and Stephen Stich (1968). “Logical Truth Revisited.” *The Journal of Philosophy* 65(17): 495–500.
- Horwich, Paul (1990/98). *Truth*. 2nd edition. Oxford: Oxford University Press.
- Jeffrey, Richard (1965/83). *The Logic of Decision*. Chicago: Chicago University Press.
- (1967/2006). *Formal Logic: Its Scope and Limits*. Ed. John Burgess. 4th edition. Indianapolis: Hackett.
- Joyce, James M. (1998). “A Nonpragmatic Vindication of Probabilism.” *Philosophy of Science* 65(4): 575–603.
- Kripke, Saul (1972/80). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, David (1973). *Counterfactuals*. Oxford: Blackwell.
- (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
- MacFarlane, John (2000). *What Does it Mean to Say that Logic is Formal?* Ph.D. thesis, The University of Pittsburgh.
- (2009). “Logical Constants.” In Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, fall 2009 edition. <http://plato.stanford.edu/archives/fall2009/entries/logical-constants/>.
- (ms/2004). “In What Sense (If Any) Is Logic Normative for Thought?” available at [johnmacfarlane.net](http://johnmacfarlane.net): presented at the Central Division APA 2004.
- Mancosu, Paulo (2010). “Fixed- versus Variable-domain Interpretations of Tarski’s Account of Logical Consequence.” *Philosophy Compass* 5(9): 745–759.

- (2011). “Explanation in Mathematics.” In Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2011/entries/mathematics-explanation/>.
- McKeon, Matthew (2004). “On the Substitutional Characterization of First-Order Logical Truth.” *History and Philosophy of Logic*, 25(3): 205–224.
- Moss, Sarah (2013). “Epistemology Formalized.” *The Philosophical Review* 122(1): 1–43.
- Peacocke, Christopher (2004). *The Realm of Reason*. Oxford: Oxford University Press.
- Pollock, John and Joseph Cruz (1986/99). *Contemporary Theories of Knowledge*. 2nd edition. Lanham, Maryland: Rowman & Littlefield.
- Prior, Arthur (1962). *Formal Logic*. Oxford: Oxford University Press.
- Quine, W. V. (1936/76). “Truth by Convention.” In *The Ways of Paradox*, revised edition, 77–106. Cambridge, MA: Harvard University Press.
- (1950/82). *Methods of Logic*. Fourth edition. Cambridge, MA: Harvard University Press.
- (1954/76). “Carnap and Logical Truth.” In *The Ways of Paradox*, revised edition, 107–133. Cambridge, MA: Harvard University Press.
- (1970/86). *Philosophy of Logic*. 2nd edition. Cambridge, MA: Harvard University Press.
- Resnik, Michael (1974). “The Frege-Hilbert Controversy.” *Philosophy and Phenomenological Research* 34(3): 386–403.
- Restall, Greg (2006). *Logic: An Introduction*. New York: Routledge.
- Rusnock, Paul and Mark Burke (2010). “Etchemendy and Bolzano on Logical Consequence.” *History and Philosophy of Logic* 31(1): 3–29.
- Russell, Bertrand and Alfred North Whitehead (1910). *Principia Mathematica, volume 1*. Cambridge: Cambridge University Press.
- Sainsbury, R. M. (1991/2001). *Logical Forms: An Introduction to Philosophical Logic*. 2nd edition. Oxford: Blackwell.
- (2002). “What Logic Should We Think With?” In Anthony O’Hear (ed.), *Logic, Thought and Language*, 1–17. Cambridge: Cambridge University Press.
- Salmon, Wesley (1963/73). *Logic*. 2nd edition. Englewood Cliffs, NJ: Prentice Hall.
- Shalkowski, Scott (2004). “Logic and Absolute Necessity.” *The Journal of Philosophy* 101(2): 55–82.
- Shapiro, Stewart (1998). “Logical Consequence: Models and Modality.” In Matthias Schirn (ed.), *Philosophy of Mathematics Today*, 131–156. Oxford University Press.
- Sher, Gila (1991). *The Bounds of Logic*. Cambridge, MA: MIT Press.
- (1996). “Did Tarski Commit ‘Tarski’s Fallacy’?” *The Journal of Symbolic Logic* 61(2): 653–686.

- Stalnaker, Robert (1968). "A Theory of Conditionals." In Nicholas Rescher (ed.), *Studies in Logical Theory*, 98–112. Oxford: Blackwell.
- (1984). *Inquiry*. Cambridge, MA: MIT Press.
- Stanley, Jason (1996). "Truth and Metatheory in Frege." *Pacific Philosophical Quarterly* 77: 45–70.
- Tappenden, Jamie (2000). "Frege on Axioms, Indirect Proof, and Independence Arguments in Geometry: Did Frege Reject Independence Arguments?" *Notre Dame Journal of Formal Logic* 41(3): 271–315.
- Tarski, Alfred (1936/1983). "On the Concept of Logical Consequence." In John Corcoran (ed.), Joseph. H. Woodger (trans.), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, 2nd edition. Indianapolis: Hackett.
- Teller, Paul (1989). *A Modern Formal Logic Primer: Sentence Logic*, volume 1. Englewood Cliffs, NJ: Prentice Hall.
- Titelbaum, Michael (2013). *Quitting Certainties*. Oxford: Oxford University Press.
- Tomassi, Paul (1999). *Logic*. London: Routledge.
- van Fraassen, Bas (1980). *The Scientific Image*. Oxford: Oxford University Press.
- Wedgwood, Ralph (2002). "Internalism Explained." *Philosophy and Phenomenological Research* 65: 349–369.