Logical Conventionalism

Jared Warren

(penultimate draft – please cite and quote from the final version, to appeal in The Oxford Handbook of Philosophy of Logic)

I. Conventionalism about Logic

It is not the case that Rudolf Carnap both is and is not a fried egg. Rudolf Carnap either is or is not a fried egg. These are truths. What is more, they are logical truths – any other sentence with the same logical form is likewise true. Also, from the fact that Rudolf Carnap was a philosopher we can conclude that either Rudolf Carnap or A.J. Ayer were philosophers. This argument is valid – truth-preserving in the strongest sense. What is more, it is logically valid – any other argument with the same logical form is likewise valid. The logical truths are also logically necessary. Their falsehood is not logically possible. And the logically valid rules preserve truth with logical necessity – it is not logically possible for all of their premises to be true while their conclusion is false. These notions – logical truth, validity, and necessity – are the central metaphysical notions of logic. They make logic metaphysically special, but they also make logic metaphysically mysterious.

Logic is also epistemologically special. In particular, logic is a priori, in at least two ways. The first way is that warrant for logical beliefs does not essentially depend upon experience. My warrant for believing that Rudolf Carnap either is or is not a fried egg is not based on any of my experiences. The second way is that logical warrant cannot be undermined by experience. You don’t have to worry about the James Webb telescope discovering that the law of non-contradiction is false. Warrant for our logical beliefs, and hence our logical knowledge, is a priori twice over. This makes logic epistemologically special, but it also makes logic epistemologically mysterious.

Ideally, our philosophical theory of logic would vindicate these features in a way that removes all mystery. Unfortunately, this is extremely difficult to do. Perhaps as a consequence, there is no consensus about the correct philosophical theory of logic. This is unsurprising; consensus in philosophy is as rare as an honest politician. Yet in the relatively recent past, there was a consensus in the philosophy of logic, at least among science-conscious philosophers. The consensus endorsed logical conventionalism.
The central thesis of logical conventionalism is simple: logic is a byproduct of our linguistic conventions. Metaphysically, this means that the roots of logical truth, validity, and necessity lie in the conventions of our language. Epistemologically, this means that warrant and knowledge in logical matters is gained using our conventions. These central conventionalist claims can be expressed in different ways. Sometimes instead of “conventions”, proponents talk of “definitions” or “linguistic rules”. And instead of being “true by convention”, logical truths are sometimes called “true in virtue of meaning” or “analytic”. Most of these differences are merely terminological. There is, at least, a strong family resemblance among the language-based theories of logic that flourished in the twentieth century. I will call all of them versions of “conventionalism”, subject to one additional constraint – conventionalist theories of logic must also lead to logical pluralism.

Any version of logical conventionalism worthy of the name appeals to language (or conceptual schemes) to found a metaphysically lightweight, epistemologically tractable, and pluralistic theory of logic.

II. Paradise Found, Paradise Lost

There are faint glimmers of logical conventionalism in the early modern era, but it is too much to credit – or blame – any pre-nineteenth century philosopher with a full conventionalist theory of logic.

Logical conventionalism partly grew out of the logicist works of Frege (1950) and Russell (1903, 1919). The logicists believed that mathematical truths were analytic, but they endorsed a loaded notion of analyticity requiring explicit definitions. This was rectified in the geometric conventionalism of Hilbert (1971) and Poincare (1914). These views took the axioms of geometry as implicit definitions of primitive geometric terms like “point” and “line”. In an explicit definition, an expression is explicitly defined using previously understood vocabulary. By contrast, an implicit definition takes the truth of certain sentences involving an expression to determine its meaning. Logical conventionalism combines the notions of analyticity and implicit definition and applies them to logic.

An important step toward this came when Russell encountered Wittgenstein. The young Wittgenstein convinced Russell that logic consisted of tautologies – trivial byproducts of our system of linguistic representation. This sounds like conventionalism, but the theory that emerged in Wittgenstein’s (1922) Tractatus was hostile to logical pluralism. According to the Tractatus,
the same logical truths emerge from every system of linguistic representation. Logic is thus tautological, but not conventional. Russell objected to some of these anti-conventionalist elements in the introduction he provided for the book Ever-after he was something like a logical conventionalist in the contemporary sense, though he never bothered to work out the details (see Russell 1959, chapter 17). Wittgenstein himself rejected the theory of the *Tractatus* less than a decade after the book’s publication.

Logical conventionalism first fully emerged in the early nineteen-thirties in the work of Wittgenstein (1974) and Carnap (1937) – see Coffa 1991 for a historical overview. At the same time, similar themes were being explored in the formal works of Gentzen (1964, 1965). The crucial move was to adapt the Hilbert-Poincare idea of implicit definition to cover logic, thus vindicating both the analytic and tautological nature of logical truth in a way that led to logical pluralism.

The standard implicit definition claim is that the truth of certain sentences determines what an expression means. In mathematics, the natural idea is that the axioms of a mathematical theory serve as implicit definitions of the theory’s mathematical terminology. In logic, this will not do. While the earliest logical systems were presented axiomatically, this was unnatural. Understanding natural language logical notions does not require endorsing an explicit axiomatic theory. Instead, the fundamental inference rules governing the uses of logical expressions like “not” and “or” and “if” and “all” are themselves taken as implicit definitions. This is an important extension, for it is not the truth of the axioms, but the validity of the rules, that implicitly defines.

This conventionalist approach became the theory of logic of the logical positivists, centered around Schlick’s Vienna circle. In the early nineteen thirties a young Oxford philosopher, A.J. Ayer, attended several of the circle’s meetings. This event was of great importance for the popularity of conventionalism, since on the basis of these experiences, Ayer wrote a popular exposition and defense of logical positivism – *Language, Truth & Logic* (second edition, 1946). Ayer’s book became one of the best-selling works of philosophy ever written. It introduced conventionalism about logic, mathematics, and all other *a priori* necessary truths to the broader public.

Of course, the west was not won for conventionalism without a fight. The metaphysicians and rationalists were not ready to go quietly into that good night. Yet most schools of philosophy that aligned with a scientific understanding of the world endorsed conventionalism, at least for logic,
and often for mathematics too, or even for necessity in general. By mid-century there was a widespread if not quite universal feeling that necessity and *a priori* had been tamed at last.

The feeling did not last. There were three main reasons for this. First, conventionalists rarely developed their theories. Wittgenstein’s views changed, Carnap moved on to other projects, and Ayer and Russell never worked out the details. Waismann (1951) and Giannoni (1971) did offer more details, but they were exceptions. Second, by contrast, critics of conventionalism developed sophisticated attacks. Many powerful criticisms sat unanswered during conventionalism’s glory days. Third, the surrounding philosophical super-structure collapsed when logical positivism was abandoned. Kripke (1980) and others undermined the equation of analytic, necessary, and *a priori* truth. Thus did the conventionalist consensus collapse into ruin. For a long while, the ruin sat unattended and abandoned. More recently, ghosts have started to stir.

Between 1980 and 2010, there were a few isolated defenses of variants of logical conventionalism (Hellman 1986) but no extended development or defense with all of the needed philosophical detail. Over the last decade I have attempted to plug these gaps, culminating in 2020 with the publication of *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism* (Warren 2020a – “SOS” below). Here I will only try to whet your appetite by saying enough to convince you that logical conventionalism is not a complete non-starter.

III. **Paradise Regained**

The section’s three sub-sections discuss the metaphysics, epistemology, and pluralism of logical conventionalism, respectively.

III.1 **The Metaphysics of Logic**

The central idea of logical conventionalism is that logical facts (facts about logical truth, validity, and necessity) are byproducts of our linguistic conventions. This divides into a number of distinct slogans, the most ubiquitous being that logical truths are true “by convention” or “in virtue of meaning”. But the early conventionalists did not do much to spin a philosophical theory out of these slogans.

In particular, there was not much clarity about the relationship between our conventions and the logical facts. What does it mean to say that logical truths are true “by” or “in virtue of” or “because of” or “owing to” our conventions? Some early conventionalists suggested that this
meant that logical truths described our conventions, as in “we use “not” according to a double negation elimination rule”. The trouble is that this makes logical truths contingent, by making them descriptions of contingent linguistic behavior. Other early conventionalists argued that conventional “truths” were themselves rules or endorsements of rules. Yet this means that logical truths aren’t actually true, since rules and endorsements themselves can’t be true. These two misfires were not what most conventionalists actually intended. They were proposed tentatively and then quickly abandoned – witness Ayer’s (1936, 1946) trajectory.

Properly understood, the key conventionalist thesis is an explanatory claim (Warren 2015a, SOS chapter 1). The logical facts, in any language, are fully explained by the linguistic conventions in said language. This is often put by saying that the conventions of a language “directly determine” the logical facts:

**Logical Conventionalism.** Logical facts in any language are fully explained by (directly determined by) the linguistic conventions of that language

This explanatory understanding is implicit in the formulations and slogans that the major twentieth-century conventionalists eventually fixed upon. It was never very precisely expressed, but Ayer (1946), Carnap (1939), Gianonni (1971), and other conventionalists eventually hit on roughly this understanding.

To say that conventions “fully” explain the logical facts is to rule out any explanatory role for non-conventional empirical information. The fact that “snow is white” is true in English is partly explained by our conventions, but also partly explained by the empirical facts. With logical truths, there is no explanatory contribution from empirical facts. To say that conventions “directly” determine the logical facts is to ensure that logical facts reflect our conventions. This closes the door on saying that, though we use the conventions of classical logic, the explanatory relationship passes through some filter ensuring that we none-the-less express only the truths of intuitionistic logic. More can be said about the needed notion of explanation (SOS section 1.III), but these clarifications are enough to make the basic idea clear.

What are linguistic conventions? There is some confusion over this, since following the influential discussion in Lewis 1969, “conventions” often refer only to social coordinating behaviors. But that is not the sense needed for logical conventionalism (SOS section 2.1). The
relevant conventions are instead the linguistic regularities that underlie our social conventions. Some early conventionalists appealed to stipulations, but this is misleading. Stipulations are metasemantically irrelevant without an underlying regularity of use. I can stipulate that “dog” means cat, but if I don’t also change my linguistic behavior, no meaning change will result (SOS, section 2.II). Historical talk of stipulations is mainly illustrative. When they were being careful about natural languages (see Carnap 1955), conventionalists instead appealed to linguistic rules and regularities. These rules sanction certain patterns of reasoning and prohibit other patterns. Carnap (1937) called them “transformation rules” in formal languages, but they clearly exist in natural languages as well (Warren 2017a, SOS chapter 2).

The early conventionalists universally agreed that these linguistic conventions could be understood in naturalistic terms. They thought this because these rules are syntactic, in a broad sense. There have been subsequent attempts to found conventionalism on semantic rules (Hellman 1986, Hansen 2021) but, among other dangers, this strategy scuttles the naturalistic credentials of traditional conventionalism. By contrast, syntactic rules concern only the physical features of sounds and marks. Of course, showing that syntactic inference rules can be naturalized requires analyzing all of the surrounding notions – rule-following, inference rule-following, inference, and belief – in scientifically respectable terms. The early conventionalists took this for granted, but the details can be filled in (Warren 2020b, forthcoming, SOS chapter 2).

Provided that linguistic conventions are scientifically respectable, logical conventionalism promises a scientifically respectable theory of logic. This promise depends on actually providing the claimed explanations of logical facts, or at least showing that such explanations exist. The conventionalist explanations lean on the popular idea that meaning is constitutively tied to language use, in a transparent fashion. In slogan form: meaning is use. There are many different metasemantic theories that would serve the conventionalist’s needs here: use theories, inferentialism, conceptual role semantics, charity or rationality-based theories, and more.

Historically, logical conventionalism was most closely tied to logical inferentialism. This is the idea that the meanings of the logical expressions – “and”, “not”, “all”, “exists”, and so on – are fully and directly determined by the inference rules according to which these expressions are used, not vice-versa.
Logical Inferentialism. The meanings of logical expression in any language are fully explained by (directly determined by) at least some of the inference rules according to which these expressions are used.

In this, the metasemantics of logical constants contrasts sharply with the metasemantics of words like “dog”, which involve causal connections to the world. Of course, logical expressions are also used in sentences that describe the world. I see Carnap and Ayer, so I say “Wow! It is Carnap and Ayer”, thus using the logical expression “and” in describing what I see. To understand this transition as involving an inference rule requires extending our notion of “inference” rule a bit (see Sellars 1968). This can be left to one side here, since narrower kinds of inference rules suffice to determine the truth-functional meanings of connectives like “and” (SOS sections 2.IV and 3.IV).

To support logical conventionalism, we only need inferentialism to apply to the logical expressions. A natural question then is: which expressions are logical? All that is strictly needed for an answer is a list of the canonical logical expressions – “and”, “not”, “all”, “or”, “if”, “exists”, and so on. But it is more satisfying to have a theoretical account of which expressions are logical. Traditional attempts to delimit the “logical” expressions appeal to features like topic neutrality and lack of empirical content. Happily, these features themselves can be understood in inferentialist terms, though I won’t go into detail about this here (SOS section 3.III).

Logical inferentialism is probably the most popular approach to the metasemantics of logic. However, most contemporary inferentialists think that restrictions must be placed on meaning-determining rules. Many of these restrictions allow, in principle, for a non-conventional element in our theory of logic, in the form of external constraints that must be met for our rules to be valid. So while some of these restricted forms of logical inferentialism lead to a muted form of conventionalism, not all of them do. By contrast, the early conventionalists were unrestricted inferentialists (Warren 2015b, SOS section 3.I-3.II). And any version of unrestricted inferentialism is a version of logical conventionalism.

Suppose you follow the rules of classical logic. If you do, then by the unrestricted inferentialist metasemantics, all of these classical rules are valid in your language. This also means that all of the logical theorems are true, necessarily so. And since there is no other explanatory input here, only the truths so explained are logically true. This same reasoning also applies to any other language. For any specific collection of rules, the details can be filled in (SOS section 4.I–4.III).
Even in broad outline though, the central point is clear. Logical truths are no more mysterious than the truth of “bachelors are unmarried”. That sentence is true because of how we use it. It is true by convention. And when considering any possible situation, bachelors remain unmarried. Likewise for logical truth and validity. Logical truths cannot be false, not because we have hit on special metaphysical facts about logical reality, but because we project our language onto any situation we describe. We do this even when considering situations at the dawn of time, or where humans never evolved. Logical truths reflect our rules of use. Since these rules are themselves syntactic rules of inference, we can call them *shadows of syntax*.

The logical truths and their necessity are byproducts of our linguistic conventions for operating with logical expressions. By the same metasemantics, only the truths so generated will have these features. In short, all and only the logical truths of classical logic will be logical truths in our language, provided our conventions are the classical ones. All of the logical facts in our language are fully and directly explained by our conventions. Logical truths are true “by” or “in virtue of” our conventions. This is the basic conventionalist story. It can be told in a number of different ways, but all of them are broadly equivalent and depend on the crucial metasemantic idea that content and truth-conditions are reflections of use. There are no legitimate questions about whether the rules we have adopted preserve truth. Meaning is determined by usage, it is not something external that our usage must answer to.

This is the very heart of logical conventionalism. Yet recently, some have called themselves “conventionalists” without tying meaning to our linguistic rules (Syverson 2002, Woods 2023) – Hattiangadi (2023: page 52) comments:

…in severing the connection between meaning and rules, these non-standard forms of conventionalism would have to eschew the sort of elegant explanation of logical validity that is provided by Warren [SOS] … In the absence of such an explanation, the connection between validity and necessary truth-preservation would be at best unexplained, and at worst, severed. A non-standard conventionalism would also give up on the equally elegant explanations of logical truth, necessity, and apriority [Ayer 1936, 1946, Carnap 1955, Warren 2020a]
Any conventionalism worthy of the name follows the positivists in tying meaning to language use in a transparent way, and then using this tie to explain the logical facts in a simple and direct fashion.

The metaphysics of logic is exhausted by our conventions. No non-conventional element plays any explanatory role in our theory of logic. We need not appeal to joints of nature, independent metaphysical facts, or anything of the sort. Logic is a byproduct of our linguistic conventions, given the way that meaning is constitutively explained in terms of language use. This, in broad brushstrokes, is the metaphysical story of logical conventionalism.

III.2 The Epistemology of Logic

Under what conditions are we warranted in believing a logical truth or in using a logical rule? Under what conditions do we have logical knowledge? These are the key questions in the epistemology of logic. Historically, the difficulty of answering these questions, along with mistaken views on the metaphysics of logic, led to rationalism and other disreputable views. One of the appeals of logical conventionalism is that it eliminates the need for anything spooky in the epistemology of logic.

Terminology is not standard, but I will call the principle notion of pure epistemology warrant. This can come in several different forms (justification, entitlement), but here they can be treated together (for my terminological choices, see Warren 2022a section 2). Logical conventionalism accounts for a priori warrant in matters of logic. The crucial idea is that there is a direct connection between warrant and meaning, and that this connection is made through the meaning-determining rules. The basic epistemic principle is very simple: speakers have automatic warrant for using the meaning-determining rules of their language. Here I will call this the meaning-epistemology-connection (MEC). This is a very plausible principle, even on its face, but we can also argue for it (Boghossian 2003).

My favorite such argument is a twist on Carnap (1937). Speaking a language is not itself an epistemic act. It is not forming a belief, it is not even adopting a belief-forming method. Instead, like riding a bike or going to a movie, choosing to speak a language is not epistemic, so it cannot be epistemically prohibited. We have a vacuous epistemic right to speak any language we please. And speaking a language is just adopting certain constitutive linguistic conventions. As such, there can be nothing epistemically improper about using a fundamental rule like modus ponens. And
warrant for using the fundamental rules extends to warrant for using the non-fundamental rules. It extends to warrant for believing anything provable from no premises, using our linguistic rules.

This argument works for any type of warrant that doesn’t require first-person cognitive access. For the remaining types of warrant, conventionalists can adapt existing access-internalist approaches (Warren 2022a Section 5).

Conventionalism also allows for logical knowledge. Suppose that “S” is true by convention in your language and you come to believe it by proving it with your language’s rules, without any undischarged premises. By unrestricted inferentialism, “S” must be true, since the rules you used are valid. And as just discussed, warrant for believing “S” is thereby generated by the MEC. One lesson of Gettier 1963 is that a warranted, true belief is not necessarily knowledge – sometimes warranted beliefs are true by accident. Accordingly, a successful analysis of knowledge probably requires an anti-accident condition. The two most popular anti-accident conditions are safety and sensitivity (see, respectively, Williamson 2000 and Nozick 1981). Let me briefly discuss how logical conventionalism guarantees that our logical beliefs are both safe and sensitive.

A belief is safe just in case it could not easily have been false. This requires that there are no relevantly similar scenarios where the belief, or a counterpart belief formed using the same belief-forming method, is false. According to conventionalism, our fundamental method for forming logical beliefs uses our linguistic conventions. In every scenario where we have these same conventions, the beliefs remain true. This is guaranteed by the inferentialist metasemantics. There are thus no relevant possibilities of error to undermine the safety of our logical beliefs. So for logical conventionalists, our logical beliefs are safe.

A belief formed by a certain method is sensitive just in case had the relevant facts been different, that method would have delivered an appropriately different belief. There are two ways for conventionalists to secure sensitivity. Which is appropriate will depend on how the requirement is understood. The first way simply notes that since conventionalism makes logical truths necessary in the strongest sense, there are no possible scenarios where the logical facts are different, so our logical beliefs are vacuously sensitive. This does not require the standard but controversial idea that all counterpossibles – counterfactuals with necessarily false antecedents – are trivial. It instead denies the coherence of supposing away analytic or conceptual truths. Even if counterpossibles sometimes make sense, counteranalytics do not.
The second way is to understand counter-*logical* conditionals in a counter-*conventional* fashion, by considering situations as described with alternative linguistic conventions (Einheuser 2006). In a situation where “$S$” is not a conventional truth in our language, we must follow different linguistic conventions. So the method of forming beliefs using *those conventions* leads to appropriate beliefs, even in this alternative situation. This reply is a bit unnatural because sensitivity seems to concerns standard counterfactuals, not counterconventionals. But perhaps conventionalists, if not other theorists, can motivate using counterconventionals here. For conventionalists, but not others, logical facts are a byproduct of our conventions. Hence it is not so strange to understand variations in the logical facts by varying the underlying conventions.

Further wrinkles could be considered, but it should already be clear that conventionalism, via unrestricted inferentialism and the MEC, allows for all needed connections between the logical facts and our beliefs. We come to believe and accept logical truths by proving them using our conventions. Nothing else is needed. In particular, we do not need any epistemic magic in the epistemology of logic, as long as we are conventionalists (see Warren 2022a for a full account).

### III. 3 Logical Pluralism

Logical conventionalism entails logical pluralism. If logical truth and validity are trivial byproducts of our conventions, and our conventions could easily have been different, then the logical truths and validities could also, *in some sense*, easily have been different.

We can describe alternative languages with different linguistic conventions. In some of these languages, the law of excluded middle, individuated syntactically, is not true. In others, the rule of *modus ponens*, individuated syntactically, is not valid. Care must be taken – it is obvious that a sentence like “Ayer is either a great writer or he is not” need not be true in every alternative language. This sentence, syntactically individuated, could even mean what “snow is black” means in English. This point threatens to make logical pluralism trivial.

In order to rescue logical pluralism from triviality, more must be said. The key point is that there are alternative languages where “not” is used as the intuitionist uses “not”, and where “not” is used as the paraconsistent dialetheist uses “not”, and so on. In such languages, the overall inferential and conceptual role of “not” is distinct from its role in English, yet it is broadly *similar*. They use “not” similarly to how we use “not”. It is similar enough in use to be similar enough in meaning – the “not” uses in these languages are *semantic counterparts*. This semantic counterpart
similarity relation (Warren 2015c, SOS section 5.III), is context-sensitive and interest-relative, like all similarity relations, yet it is of crucial importance. It is needed for characterizing both conceptual pluralism, in general, and logical pluralism, in particular. Some alternative languages have alternative logical notions, in that the semantic counterpart of our negation symbol, in such languages, doesn’t behave exactly like our negation symbol behaves in our language. Likewise for other logical notions.

This is not a problem. For the conventionalist, there is no negation up in Platonic heaven that all languages are trying to describe. Logic itself is not describing anything. There are only syntactic symbols and what is done with them, nothing else. Our logic, though extremely useful, is not written into the structure of reality like the laws of nature. In this way, logical conventionalism leads to a non-trivial version of logical pluralism that follows from the possibility of alternative logical conventions, understood with the subtlety of the semantic counterpart relation.

This pluralism makes many debates in the philosophy of logic merely verbal. If you have adopted different conventions for using “not”, then you are not automatically making a mistake when your uses of “not”-involving sentences diverge from my uses of such sentences. I can try to convince you to adopt my conventions, but I cannot convict you of error on your own terms. This point is something in the vicinity of Carnap’s (1937) famous “principle of tolerance”, but tolerance comes in different forms. There is no factual issue between proponents of different logics, but there may well be serious and important practical issues. Everyone should choose the logical tools that best suit their own interests and purposes (Warren 2018, SOS chapter 5).

Historically, talk of “conventions” was often meant to stress this pluralist aspect of the view. The term initially suggests a completely free choice, as if we choose logical conventions like we choose fruit at the grocery store. In reality though, it is neither easy nor simple to alter your fundamental conventions. Most often, linguistic conventions are neither explicitly chosen nor even explicitly represented. This point is compatible with pluralism (some conventionalists disagree – see Azzouni 2014). Even if we find it practically impossible to significantly alter our own conventions, it is easy enough to imagine aliens or alternative versions of ourselves operating with starkly different conventions. Some of these imagined beings, at least, would have a different logic than we do.

IV. Objections to Logical Conventionalism
The best way to flesh out logical conventionalism is by addressing objections and criticisms. Conventionalism has faced many objections. Even that is an understatement – conventionalism, in general, and logical conventionalism, in particular, is perhaps the single most criticized view over the last one hundred years of philosophical history. Even now, prevailing opinion seems to be that not only has the conventionalist stronghold been razed, but the wells have been poisoned and the fields salted. This is mistaken – the objections to conventionalism can all be conclusively answered.

The objections below are discussed in approximate chronological order. These discussions will, of necessity, be brief, so I also point readers to further literature on each objection. I cannot cover every objection here, but I discuss the most historically influential objections, along with two recent objections.

Quine on truth by convention (Quine 1936). The earliest major criticism of logical conventionalism is also the most historically influential. It comes from Quine.

Quine’s argument is based on the fact that there are infinitely many logical truths in our language. If these are all true by convention, then we must have made infinitely many direct stipulations. But that’s absurd. We couldn’t have made infinitely many stipulations in any finite time. Given the infinitude of logical truths, it seems the only other conventionalist option is that we stipulated general claims, rather than particular claims. For example, the claim that every instance of the law of excluded middle is true. The trouble is that we then have to move from these general stipulations to each particular logical truth, and this move requires logic. So either there is a non-conventional logical order behind our conventions, or we are off on something like the famous Carroll (1895) regress. Since these two options are exclusive and exhaustive, together this amounts to an apparently powerful argument against logical conventionalism.

Though extremely influential, this argument is powerless against plausible forms of conventionalism. Quine assumes that “conventions” can only be explicit stipulations. That is absurd. As we have already seen, conventions are instead implicit linguistic rules, principally rules of inference. When implicit conventionalism is endorsed, Quine’s argument poses no further trouble. Quine himself recognized that he was only targeting explicit conventionalism. Against implicit conventionalism he baldly claimed that implicit rules were explanatorily idle. He says this because he thinks that to say there is an implicit convention sanctioning the acceptance of a
sentence is just to say that the sentence is obvious, in a behavioral sense. However, any reasonable account of inference rule-following rebuts this charge.

For further discussion see Boghossian 1996, Azzouni 2014, Warren 2017a, and SOS sections 2.VIII and 7.III.

The master argument (Ewing 1940, Lewis 1946, Pap 1958, Lewy 1976, Yablo 1992, Boghossian 1996, Sider 2003). Another very old objection is often seen as “the” problem for truth by convention. Because of its status, I have dubbed it the “the master argument” against conventionalism (Warren 2015a). The argument is sometimes presented using propositions, at other times using states of affairs or facts. Here is one fact-based version of the argument.

Logical conventionalism claims that our conventions make logical truths true. But sentences are true just in case they express a fact. So we have something like the following general equivalence principle:

\[
\text{sentence } "S" \text{ is true just in case there is a fact } f \text{ and } "S" \text{ expresses } f
\]

From this it is argued that if conventions make it the case that “S” is true, then they must also – by this equivalence – make it the case that there is a fact f that “S” expresses. Yet this seems absurd. Our conventions concern language, as the left-hand-side of the equivalence does, while the right-hand-side is instead about the world. Our linguistic conventions cannot generate non-linguistic facts. Hence, the argument concludes, truth by convention is a non-starter.

Despite its popularity, the argument suffers from several crippling failings. For one thing, “making true” is not the best understanding of the crucial conventionalist relation. It sounds either causal or metaphysical, while – as discussed above – the conventionalist claim is explanatory. And explanatory contexts are hyperintensional, that is, substitution of necessarily equivalent terms into these contexts can result in sentences with different truth values. So the equivalence above, even if necessary, does not allow us to reason from our conventions determining that “S” is true to our conventions determining that there is a fact f. In short, the argument commits a hyperintensional fallacy. This point alone completely sinks the argument.

Questions remain about how conventionalists should understand logical facts. The answer depends on what is meant by “fact”. If “logical facts” are supposed to be objective parts of some
metaphysical reality that logic is aiming to describe, then conventionalists reject logical facts altogether and so reject the equivalence above. If “logical facts” are instead supposed to be lightweight projections of language, then the conventionalist can accept the above equivalence, and even, if so inclined, allow that our conventions “make” such facts. But this move risks confusion over the modal and temporal status of logical truths, so it is often better avoided, especially since lightweight facts play no load-bearing role in logical epistemology. The central point is that there is no role for “fact”-talk that requires a non-conventional theory of logic. This is all that conventionalists are committed to with respect to “logical facts”, qua conventionalists.

And standard “fact” talk is not philosophically loaded. We often use it when making practice-internal explanatory claims. As an illustration of this, recall that in the Harry Potter series it is true that Harry was invited to the Hogwarts school of witchcraft and wizardry. What makes this true? Well, in one sense, it is true because J.K. Rowling wrote it that way. But if a student were to give this answer on a literary comprehension examination, we would not give them credit. To get proper credit, they would need to cite other facts within the stories. These are further facts on the same explanatory level as the fact to be explained. Likewise, a logical conventionalist might appeal to other logical facts to explain given logical facts. But this type of explanation does not assume an alternative philosophical account of logic. It is instead a part of our ordinary explanatory practices themselves. A part that accounts for the existence of some of the intuitions that proponents of the master argument attempt to leverage against conventionalism. Our explanatory practices have many different strands. When we recognize this, we see that the master argument fails, despite its influence.


The categoricity problem (Carnap 1943). Inferentialists say that inference rules determine the meanings of logical expressions. Yet, for classical logic, this does not seem true, at least for standard inference rules.

The basic problem was first noticed by Carnap. The details are technical, but in brief, the standard validity of standard natural deduction rules does not strictly force standard meanings on the logical expressions. So any form of inferentialism wedded to the standard assumptions has trouble accounting for classical logic.
There are several formal proposals that avoid this defect, the first was offered by Carnap himself. What is really needed is a response that philosophically aligns with inferentialism. I think the best approach recognizes that inferences can crucially involve rejection as well as acceptance. This is lost in standard natural deduction formalisms, but is rectified in bilateralist natural deduction systems – introduced in Smiley 1996. Bilateralist systems solve Carnap’s problem for the sentential connectives.

The problem for quantifiers is not resolved by bilateralism, but by the open-endedness of our logical rules. Logical rules like modus ponens are open-ended – they continue to apply as we add new sentences to our language (McGee 2000). I think that both bilateralism and open-endedness are independently needed to understand our inferential practices, and happily, together they also resolve Carnap’s categoricity problem.

For further discussion see Carnap 1943, SOS section 3.IV, chapter 10, and Murzi and Topey 2021.

The analytic-synthetic distinction (Quine 1951). Historical conventionalists often freely moved between saying that logical truths were “true by convention” and saying that they were “analytic” or “true in virtue of meaning”. Yet, starting in mid-century, launched influential arguments against the cogency of the analytic-synthetic distinction. To this day, there is disagreement about the exact nature of his criticisms. Some think he was rejecting the distinction entirely; others think he was arguing that the distinction was coherent but vacuous because nothing fell completely on the analytic side of the divide.

The canonical reply to both readings is from Grice and Strawson (1956) – there is far too much agreement over cases, even novel cases, for it to be plausible that the distinction is either incoherent or vacuous. Speakers agree about the analyticity, or not, of the vast majority of cases. And even if there are borderline cases, this is not a reason to deny the existence of clear cases. There are borderline cases of “bald”, but there are also both clearly bald and clearly not-bald people.

Perhaps the deeper worry suggested by Quine concerns the explanatory role played by analyticity, but this matter cannot be judged in advance. We can only assess theories that appeal to analytic or conventional truths on a case-by-case basis. To do otherwise is to rule against conventionalism before it has had its day in court.

For further discussion see Grice and Strawson 1956, Boghossian 1996, and SOS section 13.I.
Universal revisability (Quine 1951, Williamson 2007). Quine also claimed that our attitude toward any sentence can change on the basis of experience. This puts apparent pressure on the idea of conventional truths as sentences that must be accepted in every empirical situation.

This seems like a problem, but actually, universal revisability is perfectly compatible with conventionalism. In fact, Carnap (1937) explicitly embraced the universal revisability of all sentences long before Quine’s challenge. Conventionalists disagree with Quine only in thinking that some such revisions amount to a move to a new language, while others do not. This is not a denial of universal revisability, it is only a denial of universal revisability without meaning change.

In Quine’s argument – and in modern variations like Williamson 2007 – it is tacitly assumed that we can always translate homophonically across revisions – syntactically individuated type sentences can be translated to the same type sentences. But this assumption is false. To think otherwise is to commit what I have called “the translation mistake”, the mistake of thinking a homophonic translation is appropriate when it is not.

Quine himself later (1970) argued against homophonic translation in cases of logical deviance. There is debate about whether this is consistent with his earlier arguments. For most of us, sentences whose truth must be respected by any coherent translation count as analytic.

For further discussion see Carnap 1963, Burgess 2004, Warren 2018, 2021, 2022a section 6, and SOS section 7.V.

The basic-to-derivative objection (Dummett 1959). From a reading of Wittgenstein’s (1956) late writings on mathematics, Dummett extracted a challenge that directly targets implicit conventionalism. This challenge pops up in many related, but not strictly identical, forms (Soames 2003 chapter 12). I will discuss one central version of the challenge.

Suppose that the rules of “and” elimination and “and” introduction are our basic, meaning-determining conventions for “and”:

(“and” introduction) From $\phi$ and $\psi$ infer $\Gamma \phi$ and $\psi \setminus$

(“and” elimination) From $\Gamma \phi$ and $\psi \setminus$ infer $\phi$, from $\Gamma \phi$ and $\psi \setminus$ infer $\psi$
Many other rules for sentences containing only “and” essentially are then forced to be valid. For instance, “and” is symmetric – from $\Gamma \phi$ and $\psi \vdash$ you can infer $\Gamma \psi$ and $\phi \vdash$ and vice-versa. By assumption, this symmetry rule is valid but not basic. However, it can easily be derived from the basic rules. In this way, we can argue that, if the basic rules for “and” are valid, and by inferentialist principles, they are, then the derivative rules for “and” are also valid. The critic seizes on this point.

The critic’s claim is that the connection between basic and derivative conventions is itself a logical connection that must be in place apart from our conventions. That is, they argue that the conventionalist is here assuming a non-conventional logical connection governing relationships between our conventions. In SOS I called these arguments “variations on a theme by Quine” because of their debt to Quine 1936. Yet unlike Quine’s argument, these variations directly target implicit conventionalism. Some of our implicitly followed rules will be basic, and others derivative. This feature makes the challenge novel and, if cogent, more difficult to answer than Quine’s objection.

Fortunately, the objection is mistaken. The dispositions and behaviors that constitute directly following the basic rules also, themselves, constitute indirectly following the derivative rules. A derivation of a non-basic rule from basic rules brings out this relationship, but it is the height of silliness to act as if this situation involves an appeal to some super-logical connection between our conventions. Instead, we simply have a logical practice where following certain rules directly constitutes following other rules indirectly. The practice is one and the same.

There is nothing but language use behind our logic, though of course we describe various connections and explanatory points using our language. Is there some other option?

For further discussion see Warren 2017a and SOS section 7.IV.

The tonk objection (Prior 1960). The most famous objection to unrestricted inferentialism is Prior’s “tonk” objection. Prior was arguing against the claim that any pair of introduction and elimination rules could implicitly define a meaningful logical connective. His counterexample is “tonk”, which is defined by the following rules:

("tonk" introduction) From $\phi$ infer $\Gamma \phi \ tonk \psi \ \vdash$

("tonk" elimination) From $\Gamma \phi \ tonk \psi \ \vdash$ infer $\psi$
Putting these rules together with the normal structural features of deduction allows us to prove any claim from any other claim. If our language has at least one theorem, this means that every sentence in the language is a theorem. Prior’s example seems to show that we need constraints on which rules can define meaningful connectives.

All previous inferentialists have agreed, mainly on the strength of this one example, that we must move from unrestricted to restricted inferentialism. That is, they have looked for some restriction on which collections of rules can found meanings. There are many proposals for this – consistency, conservativeness, harmony, and more. Yet a restriction is not actually required.

If language users really and truly follow the tonk rules, they thereby alter the usage patterns in their language so much that every sentence is now akin to a logical or conceptual truth. In a tonk language of this kind, no empirical information about the world can be expressed. The tonk language is totally useless for most purposes. This makes it a silly language. We say this, from the outside. Yet we can and should still say that, in this tonk language, the tonk rules are valid. Of course, the tonk rules are invalid in English – from “2 + 2 = 4” we get a tonk-proof of “Saul Kripke is Plato” – but this does not show that they are invalid in a tonk language.

It is not easy to actually follow the tonk rules the way we follow modus ponens. Doing so would swamp all of the crucial connections our language has set up with the world. And if our language also includes rejection rules, following the tonk rules is impossible, full-stop. You cannot have the dispositions that constitute following the tonk rules while at the same time following our other rules. This means that beings like us can’t speak certain languages, even in principle. But nor can we both catch and miss a bus. Recognizing this kind of impossibility does not involve any restriction of inferentialism.

For further discussion see Warren 2015b, Warren 2022a section 5, and SOS section 5.II.

The contingency of conventions (Blackburn 1987). In language, we can consider what would have happened had things gone differently. We can even talk and think and reason about situations where humans, or their linguistic conventions, never came to be.

Some critics have tried to raise an objection here. Logical truth is necessary, yet our conventions are contingent. How, they ask, can we account for necessary truths in contingent terms? Instances of the law of non-contradiction are true in our language, necessarily so. Yet suppose we spoke a language in which those same sentences were false. Aren’t conventionalists
forced into saying that the law of non-contradiction is false in that situation and hence contingent? No; for the imagined situation is not a situation in which the law of non-contradiction is false. It is merely a situation with a language in which syntactic contradictions and their negations do not express what they express in our language.

How do we explain logical necessity using contingent conventions? Quite simply; the necessary truths are those sentences that don’t vary their truth values in counterfactual contexts. This itself is part of how we use language. Logical rules are not built into any factual situation we are describing. They are instead part of our descriptive resources, here, for describing alternative situations. Some descriptions are logically incoherent, by our rules, hence, those supposed situations are logically impossible. We import our logical apparatus into any counterfactual situation. Thus, for conventionalists, we thereby import all of our logical truths too. This is all that there is to the metaphysical necessity of logical truth.

In these ways, our logic is projected onto counterfactual scenarios of all kinds. Projected onto, not found within. Metaphysical accounts of the necessity of logic are useless pseudo-explanations. Exactly parallel questions can be raised about the eternality of logic, to which the conventionalist gives exactly parallel answers – logic is eternal because logical rules are insensitive to distinctions of tense and time.

For further discussion see Sidelle 2009, Warren 2022b, and SOS sections 4.I and 7.I

The harmony objection (Dummett 1991, Nyseth forthcoming). Another canonical objection to unrestricted inferentialism is based on harmony. Harmony is the requirement that the elimination rules for a logical constant must harmonize with the introduction rules for the constant. In its strongest form, this allows only the introduction rules for a connective to be meaning-determining, because introduction rules uniquely determine harmonious elimination rules.

Yet if we take seriously the idea that meaning is determined by use, harmony constraints are implausible. Suppose that a language has a binary connective C that speakers of the language introduce using a rule that is structurally identical to both disjunction introduction and tonk introduction. Whether this connective means disjunction or tonk depends on which elimination rules are paired with these introduction rules. And these different meanings correspond to different patterns of use. The harmony approach denies this tie between meaning and use, and with it, denies the metasemantic basis of inferentialism.
With tonk, the elimination rule is intuitively *too strong* for the introduction rule. This kind of harmony failure was addressed above. The remaining case is when the elimination rule is intuitively *too weak* for the introduction rule. Consider the following rule:

(“tink” introduction) From \( \phi \) and \( \psi \) infer \( \phi \upharpoonright \phi \uparrow \psi \uparrow \)

(“tink” elimination) Infer \( \chi \) from \( \phi \uparrow \phi \uparrow \psi \uparrow \) when \( \chi \) has been inferred from \( \phi \) and also from \( \psi \)

This connective has the introduction rule of conjunction but the elimination rule of disjunction. The harmony fan argues that, given the tink introduction rule, both \( \phi \) and \( \psi \) must be properly inferable from \( \phi \uparrow \phi \uparrow \psi \uparrow \), for – the harmony fan asks – how could it be otherwise? Hence, tink is just conjunction.

From an unrestricted inferentialist perspective, this view is strange. If the tinkers really and truly *decline* these inferences, with their heads clear and their eyes wide-open, we should interpret them in a way that vindicates this usage. This means that we can’t translate “tink” as a normal, time-stable truth-function, but so what? We can instead use whatever exotic devices are needed to make sense of the tink language – hyperintensional operators, context sensitivity, utterance relativity, or anything else in the semantic menagerie. Unrestricted inferentialism is the only approach that truly accepts that meaning is determined by use. There is no additional, practice-transcendent feature that determines meanings. Say it with me: *meaning is use*.

For further discussion see Burgess 2005, Warren 2015b, and SOS section 5.1-II.

**Meaning-determining conventions** *(Fodor and Lepore 1992, Rescorla forthcoming)*. A natural question for any inferentialist theory is: which of our conventions determine meanings? So-called *holists* claim that all rules we use for an expression are (partly) meaning-determining for the expression. So the meaning of “and” is determined, in part, by every valid inferential transition that involves “and” over our entire language. The nice thing about holism is that we don’t need any distinction between the rules that are meaning-determining and the rules that are not. The not-so-nice thing about holism is that it entails that *any* divergence in use results in a difference in meaning.
Accordingly, most inferentialists and conceptual role theorists are non-holists, including most conventionalists. But non-holist inferentialists owe us a story about which of our rules are meaning-determining. This is often run together with questions of analyticity, but the issues are distinct – a sentence like “if Ayer is a great writer, then Ayer is a great writer and Ayer is a great writer” is analytic, but it is not meaning-determining by almost anyone’s lights.

The obvious non-holist approach appeals to the distinction between basic and derivative patterns of use that has already been discussed above. Usually this distinction can be understood in a straightforward counterfactual fashion – use pattern 1 would not be in place without use pattern 2, but not vice-versa. However it is understood, some of our patterns of use are explanatorily prior to others. The patterns of use that sit at the bottom of these explanatory dependence chains are the basic patterns of use. When the basic patterns are inference rules, call them basic rules. There may be some indeterminacy about which rules are basic in natural languages, but it is very difficult to deny the distinction entirely. The natural view is that the meaning-determining usage patterns are the basic usage patterns, and the meaning-determining rules are the basic rules.


Conventionalism and disquotational truth (Field forthcoming). Field has argued that conventional truth does not disquote – we cannot automatically conclude that $S$, even if we have established that “$S$” is conventionally true. So, while logical conventionalists can establish that “Ayer is either a great writer or he is not” is conventionally true, they cannot use this to establish that Ayer is either a great writer or he is not. If this is right, then “conventional truth” is not truth. Conventionalists have simply mislabeled the notion of following from our conventions as “truth”.

Note though that conventionalists don’t actually try to prove logical truths in their home language by first establishing a conventional truth claim and then disquoting. Instead they prove the sentences themselves from no premises using their conventions and accept them on that basis. Their truth then trivially follows, using our language’s own disquotational truth predicate. In similar ways, we conventionalists can prove that all theorems of our language are true, and that anything provable from truths using the rules of our language is also true, and so on (see SOS, chapter 6). This reasoning is accepted by Field and everyone else, yet unlike Field and everyone else, conventionalists have a metasemantics that answers challenges about supposed
misalignments where we prove something that is not actually true – for inferentialists, such misalignments are impossible.

Field’s objection implicitly denies this conventionalist metasemantics. If our basic conventions are automatically valid, then any sentence provable from no premises using only our basic conventions is thereby both true and necessary. This is a constitutive principle of metasemantics. To deny the principle is, thereby, to reject conventionalism. So any argument against conventionalism that rests on such a denial is question-begging.

The defining metasemantic principles of conventionalism were stated (in section III.1) using a “true in $L$” predicate, for variable language $L$, rather than our own language’s disquotational truth predicate. This was done because the principles apply to all languages, not just our own. However, the principles could have been expressed using a disquotational truth predicate, at the cost of some unnaturalness and a few added normative claims about translation into our language. Obviously, when “$S$” is true in $L$, we cannot automatically disquote and accept “$S$” in our home language, but this is a feature, not a bug. If “$S$” is true in $L$, then the translation of “$S$” into my language is also true, and when my language is $L$, this translation is homophonic, allowing direct disquotation. And the inferentialist principles also entail that the theorems of my language are true in my language, and this de se truth predicate will either directly disquote, or will be provably equivalent to my language’s disquotational truth predicate. Either way, there is no trouble in moving from conventionalist principles and truth-claims about our language to object-level claims within our language, so Field’s objection fails even when taken on its own terms.

Admissible rules and implicit conventionalism (Golan forthcoming). Golan has argued that there is a hidden problem with implicit conventionalism’s treatment of validity. As we have seen, according to implicit conventionalism, the validity of basic logical rules explains the validity of the derivable but non-basic rules. These derivative rules are indirectly valid by convention, while the basic rules are directly valid by convention. The trouble is supposed to come from admissible but non-derivable rules of inference.

The notion of “admissibility” holds (roughly) that a rule is admissible just in case, when all instances of the premises of the rule hold, so too does the relevant instance of the conclusion of the rule. Yet for some formal frameworks for some logics, there will be admissible rules that are not derivable. Golan sees this as a problem for conventionalists because conventionalists should,
he argues, see some of these admissible but non-derivable rules as likewise valid by convention, but since they are not derivable, there is no conventionalist account of their validity.

Golan poses his challenge using a formal framework, known as a sequent calculus, in which both the “premises” and “conclusions” of the “rules” are themselves entailment claims. These rules state facts about our conventions and the entailment claims that constitute them have conventional witnesses. As such, these are not the kinds of “rules” that implicit conventionalists were discussing in their treatments of basic and derived rules. And conventionalists have already explicitly recognized that the validity of sequent-style rules can be explained by general features of language use instead of derivability (SOS sections 3.II and 7.IV), so admissible but non-derivable sequent rules are not a serious problem.

What is more, even for standard inference rules, conventionalists from Carnap (1937) on have always allowed that conventional factors other than derivability can explain a rule’s validity (SOS section 11.I). So the challenge is both misconceived and misaimed.

V. The Return of Logical Conventionalism

Seventy-five years ago logical conventionalism was the most widely accepted philosophical theory of logic. It didn’t attain this status through authority. It attained it because, in that age of scientific miracles, with metaphysical speculations having temporarily exhausted themselves, philosophers and non-philosophers alike yearned for a scientifically plausible theory of logic. Conventionalism was that theory. As argued here, when suitably updated, conventionalism still is that theory. In fact, I don’t think it has any serious rival as a comprehensive philosophical theory of logic.

Philosophy is currently in the middle of a scholastic, metaphysical bubble. Like all such bubbles, this one too will eventually burst. So, an advertisement to all philosophers and serious thinkers: if you want a plausible, metaphysically sane, epistemologically sensible, and elegant theory of logic then conventionalism is for you. The confusions have been cleared up, the objections have been answered, and the details have been filled in. Logic is a priori, analytic, and necessary. We can have all of this, while keeping our metaphysical and epistemological feet on the ground. When it comes to logic, we can eat our cake and have it too. Conventionalism is how:

Bibliography


