A popular view in the philosophy of logic is inferentialism — the view that the meanings of logical constants like “and” (\(\land\)), “if” (\(\rightarrow\)), and “not” (\(\neg\)) are fully determined and explained by the inference rules according to which they are used. This view naturally combines with general metasemantic theories that explain linguistic meaning in terms of language use (sometimes called inferential role or conceptual role semantics). The simplest inferentialist view claims both that (i) any collection of rules for an expression can determine a meaning for the expression and (ii) meaning determining rules are automatically valid (necessarily truth preserving).

But the simple, unrestricted, version of inferentialism has been universally rejected by philosophers because of putative counterexamples like A.N. Prior’s tonk connective: for any sentences \(p\) and \(q\), tonk’s introduction rule allows us to conclude that \(p_{tonk}q\) from \(p\), and tonk’s elimination rule allows us to conclude that \(q\) from \(p_{tonk} q\). So with the tonk rules in hand, we can conclude that Saul Kripke was born before Plato from \(2 + 2 = 4\); but while \(2 + 2 = 4\), Saul Kripke was not born before Plato, so the tonk rules are clearly and obviously invalid, or so it is universally thought. I will argue that this consensus is mistaken.

There are two standard responses to tonk: (1) give up on the idea that any collection of rules can determine a meaning; (2) give up on the idea that meaning determining rules are automatically valid. But I think we can accommodate tonk without giving up on either of these ideas. After explaining inferentialism, I’ll argue for this unpopular and controversial claim by considering the possibility of a tonk language (Tonklish) with tonk speakers (Tonkers).

1. Inferentialism

Metasemantics concerns the nature of semantic properties like truth, reference, and meaning. One popular family of metasemantic theories explains semantic properties in terms of language use; call any theory in this family a use-based metasemantics. Among use-based metasemantics are theories that explain semantic properties in terms of inference rules; call such theories versions of inferentialism. Inferentialism comes in a variety of forms — local inferentialism.

---

applies inferentialism to some particular branch of discourse; *global inferentialism* extends inferentialism to all branches of discourse.

The central idea of inferentialism is that linguistic meaning is determined by rules of language use; we can sum this up in the following principle:

**Meaning Inference Connection (MIC):** the meaning of any linguistic expression is fully determined by the rules according to which it is used.

To flesh out the MIC, we need to understand patterns of language use, the nature of linguistic rules, linguistic rule-following, and the relationship between them.²

In neutral terms we can admit that any linguistic item ⊩ (word, sentence, clause, etc.), in language community C, is *used* in some manner where *use* includes actual tokenings in addition to unrealized dispositions to token the item in various counterfactual circumstances. Let’s call this ⊩’s *pattern of use* in C. So the pattern of use of the sentence “it is raining” in our language community includes not just the actual uses of this sentence, but also speakers’ dispositions to use the sentence in unrealized circumstances where, for example, water is falling from the sky in their location.

The picture of language lurking behind the MIC holds that as far as the information given in ⊩’s pattern of use isn’t just semantically irrelevant noise, it is explained by the members of community C following certain rules of use when using ⊩. By a “rule of use” I mean not just inference rules, narrowly construed, which tell us to accept some sentence (the rule’s conclusion) whenever we accept some other sentences (the rule’s premises), but also other rules that tell us to expand or contract the set of sentences we accept or reject in certain circumstances. Call this the broad understanding of inference and understand ‘inference rule’ in the MIC in accordance with this broad understanding.

Expression ⊩ is used according to rules R₁, R₂, . . . , Rₙ when speakers in the relevant language follow these rules when using ⊩ and, in a sense to be explained, these are the only rules they follow when using ⊩. By this I mean that the rules R₁, R₂, . . . , Rₙ fully explain ⊩’s total pattern of use. This doesn’t rule out the existence of various derivative rules according to which the speakers in the community use ⊩, but it does demand that following the rules R₁, R₂, . . . , Rₙ suffices for following these derivative rules. This means that the inferentialist is not automatically committed to a problematic form of holism, where every rule for an expression in our language is meaning determining for that expression.³

As long as the meaning determining rules — the rules according to which expression ⊩ is used — suffice for generating the total pattern of use, only they need be meaning constituting.

To illustrate these points, consider the conditional, “if...then...”/“→”. Since their invention by Gentzen and Jaśkowski in the 1930s, philosophers and logicians have been attracted to the idea that natural deduction rules in formal systems helpfully model (some of) the inference rules employed in natural languages like English.⁴ In such systems, the typical introduction and elimination rules for the conditional are (CP) — conditional proof or →-introduction; and (MP) — modus ponens or →-elimination:⁵

\[
\begin{align*}
\text{(CP)} & \quad \frac{\psi}{\phi \rightarrow \psi} \\
\text{(MP)} & \quad \frac{\phi \rightarrow \psi \quad \phi}{\psi}
\end{align*}
\]

In these systems, there are other valid rules for the conditional, such as (HS) — hypothetical syllogism:

\[
\begin{align*}
\text{(HS)} & \quad \frac{\phi \rightarrow \psi \quad \psi \rightarrow \chi}{\phi \rightarrow \chi}
\end{align*}
\]

---

2. It is also worth noting that I am thinking of the determination relation as a fairly direct explanatory relationship.


4. See Gentzen (1934) and Jaśkowski (1934).

5. In (CP) the brackets indicate that the bracketed formula is an assumption; an assumption with a numeral superscript n is discharged at the line indexed by n.
However, \( (HS) \) and other similar rules don’t need to be included when setting up the system, since given the rules \( (CP) \) and \( (MP) \) and standard structural assumptions, the rule \( (HS) \) is provable as a derived rule. In essence, following \( (CP) \) and \( (MP) \) suffices for following \( (HS) \): giving up a derived rule like \( (HS) \) necessitates giving up a primitive rule like \( (CP) \) or \( (MP) \).

In a formal system, it’s easy to distinguish the primitive rules from the derived rules, since we are simply given a list of primitive rules when being taught the system. In a natural language, however — which is what we’re ultimately concerned with — things aren’t so neat and tidy. Despite this, there are ways to recapture something like the distinction between primitive and derived rules. We could, for example, distinguish between those inferential transitions that speakers of a language accept directly and those that they accept only because they can be seen to be valid using directly accepted inferences. Let’s call the directly accepted inferences direct rules and the indirectly accepted inferences indirect rules; it’s natural to think that only the direct rules for an expression \( \oplus \) are the rules according to which \( \oplus \) is used, in the sense of the MIC. The standard natural deduction introduction and elimination rules for a connective are important because they can often plausibly be taken to be direct, as in the case of \( (CP) \) and \( (MP) \). Speakers accept natural language versions of these rules not because they have convinced themselves of their validity using other rules, but simply because they do. The only constraint on direct rules that I have introduced is this: taken together, they need to cover and explain the entire pattern of use for the language.

Thus far I have helped myself to the notion of following a linguistic rule, but rule-following is a notoriously controversial matter and has been at least since Kripkenstein’s influential discussion. I won’t be offering any analysis here; still, it’s worth saying a bit to flesh out what I am and am not assuming about rule-following: when I say that an agent \( S \) follows rule \( R \) I mean not just that \( S \) is disposed to act in accordance with \( R \) in the vast majority of some class of independently specified standard situations, but also that \( S \) is disposed to accept corrections toward \( R \) quite readily; that we can expect \( S \)’s behavior to continue in this manner in a wide variety of hitherto unobserved situations; that there is some positive assessment by \( S \) of behavior in conformity with \( R \) and some negative assessment of behavior not in conformity with \( R \), etc.

Crucially, I don’t think that rule-following requires an explicit representation or recognition of the rule being followed, for that seems to require too much. In general, I think that some complex dispositionalist account of rule-following is correct. Although Kripkenstein argued at length against dispositionalism, it remains the default view of rule-following amongst naturalists.

Before proceeding, three further points of clarification concerning the MIC are worth making: (i) in actual cases there will often be a large amount of indeterminacy about exactly which rules a community is following, but this is to be expected and shouldn’t seriously trouble us; (ii) I’ll generally be assuming that every member of a given linguistic community follows the same linguistic rules; this is a simplification that prescinds from issues concerning coordination between speakers; (iii) The MIC should not be confused with its converse, which says that the meaning of an expression determines which rules of use for it are followed. A natural reading of this would entail that any difference in rules followed leads to a difference in meaning and this is something that inferentialists differ on. I personally find it implausible, but in any case, I won’t be assuming it here and so the MIC shouldn’t be read as entailing its converse.

With these clarifications in place, we can see that there is a sense in which everybody — or, at least, every naturalist — accepts something like the MIC understood as involving inference in the broad sense. If the linguistic rules underlying our patterns of language use don’t determine linguistic meaning, then it isn’t clear what else possibly could. For this reason, it would be uninformative to call anyone who accepted the MIC an “inferentialist”. The distinctive
thesis of inferentialism for truth-apt, assertoric branches of discourse, is that rules determine meaning by being transparently valid:

**Meaning Validity Connection (MVC):** The meaning determining rules for any expression are automatically valid (necessarily truth preserving)

When the meaning of expression \( \oplus \) is determined by the rules of use \( R_1, R_2, \ldots, R_n \), say that the rules \( R_1, R_2, \ldots, R_n \) implicitly define \( \oplus \) and that the rules \( R_1, R_2, \ldots, R_n \) are meaning constituting for \( \oplus \). The import of the MVC is that if rule \( R \) is meaning constituting, then \( R \) is thereby valid, i.e., \( R \)'s validity is wholly and fully explained by the fact that it is meaning constituting.\(^{12}\) In essence, the MIC and the MVC together tell us that the meaning of an expression \( \oplus \) is whatever it has to be in order for \( \oplus \)'s meaning constituting rules to be valid. Call anyone who endorses both the MIC and the MVC for the expressions in some branch of discourse \( X \) (e.g., logic, mathematics, etc.) an inferentialist about \( X \).

Inferentialism was first endorsed by the Wittgenstein of *Philosophical Grammar* and the Carnap of *The Logical Syntax of Language*.\(^{13}\) Here’s Wittgenstein on negation:

> There cannot be a question of whether these or other rules are the correct ones for the use of “not” (that is, whether they accord with its meaning). For without these rules the word has as yet no meaning; and if we change the rules, it now has another meaning (or none), and in that case we may just as well change the word too.\(^{14}\)

And here is Carnap on the general reversal being pulled off:

> Up to now, in constructing a language, the procedure has usually been, first to assign a meaning to the fundamental mathematico-logico symbols, and then to consider what sentences and inferences are seen to be logically correct in accordance with this meaning. . . . the connection will only become clear when approached from the opposite direction: let any postulates and any rules of inference be chosen arbitrarily; then this choice, whatever it may be, will determine what meaning is to be assigned to the fundamental logical symbols.\(^{15}\)

Many philosophers have been attracted to inferentialism, at least in certain domains of discourse. Inferentialism is extremely popular in the philosophy of logic, where it has been endorsed in some form by Paul Boghossian, Michael Dummett, Ian Hacking, and Christopher Peacocke, among many others.\(^{16}\) Logical inferentialists think that the direct rules for our logical expressions, perhaps including natural language analogs of the rules \( CP \) and \( MP \), are automatically valid and any doubts about this are empty. If you accept the MIC and the MVC there is simply no further question about whether the direct rules are valid for the meaning of the symbol “\( \rightarrow \)”, for these rules determine the meaning of this symbol in our language.\(^{17}\)

A substantive difference between Wittgenstein and Carnap is visible in the quotes above, viz., Wittgenstein suggests that rules can fail to determine a meaning for an expression, while Carnap suggests that any rules, chosen arbitrarily, can implicitly define a meaningful expression.\(^{18}\) We can sum up the Carnapian position in the following principle:

**Meanings Rules Connection (MRC):** Any collection of inference rules used for and involving an expression can be meaning constituting for that expression

---

12. When the inference is understood broadly, validity will also need to be understood broadly, e.g., if we have a meaning constituting rule telling us to reject \( q \) whenever we accept all of \( p_1, \ldots, p_n \), then the rule will be valid just in case the conclusion is false whenever the premises are all true. See Smiley (1996) and Rumfitt (2000) for relevant discussion.
13. Other early endorsements or partial anticipations are found in Gentzen (1934) and Popper (1947).
15. Carnap (1934), page xv.
17. As Carnap (1943) was the first to realize, standard natural deduction systems for classical logic fail to determine standard meanings even for all of the connectives of sentential logic; this can be remedied by adding rejection rules or by working in a multiple conclusion system.
The MRC is, at least at first-glance, a fairly natural principle. It sums up the intuitive idea that meanings (or concepts) are cheap. Call anyone who endorses the MIC, MVC, and MRC in some branch of discourse X, an unrestricted inferentialist about X.

It’s important to note that MRC says that any rules “used for” and “involving” an expression can determine a meaning for that expression. This differs from the letter of Carnap’s claim concerning choosing rules of inferences arbitrarily, but it arguably captures the intended metasemantic import of Carnap’s remark. It wouldn’t make sense to say that rules $R_1, R_2, \ldots, R_n$ determine a meaning for expression $\phi$ if $\phi$ didn’t occur in any of the rules (or in any instances of the rules if they’re schematic). And if the rules couldn’t coherently be used by language users of any kind, the inferentialist wouldn’t be forced into saying that those “linguistic rules” aren’t linguistic rules at all, since no possible language users could use them (see sections 5 and 6 below for further relevant discussion). This is just to say that the MRC concerns only rules that can sensibly be deemed linguistic rules: global inferentialists aren’t forced into saying that the castling rule in chess determines a meaning for “cat”. It is only the rules that compress and distill a possible pattern of use for an expression that determine a meaning for the expression; what the MRC adds to the MIC is a simple closing off of the possibility that the direct rules for an expression fail to determine a meaning for the expression.

Unrestricted inferentialism offers the hope of explaining linguistic meaning in terms of rule-governed patterns of actual and possible language use, eliminating the need for abstract objects like meanings or propositions as load-bearing features in metasemantics. As such, the view is extremely attractive on naturalistic grounds. But despite its obvious virtues, unrestricted inferentialism has been — as far as I am aware — rejected by every single philosopher post-Carnap. It has been rejected because of purported counterexamples; the most pressing and famous of which is Prior’s tonk connective.

2. The Challenge of Tonk

Arthur Prior’s infamous tonk rules pose a challenge to unrestricted inferentialism:19

$$\begin{align*}
\text{Tonk Introduction:} &\quad \phi \\
\text{Tonk Elimination:} &\quad \phi \text{tonk } \psi
\end{align*}$$

To more effectively discuss the challenge that tonk poses to inferentialism we need to introduce some terminology. In logic textbooks, a language consists of a lexicon plus grammatical rules determining which sequences of items drawn from the lexicon are sentences. Formal systems of various sorts are often used as toy models of messy natural languages like English (we saw an example of this above with the discussion of natural deduction rules), but the formal systems that model natural languages aren’t “languages” in the logician’s sense, but rather languages plus a proof theory (including structural rules of deducibility). Say that a full language is the combination of a language in the standard formal sense together with a proof theory; a full language is supposed to model a natural language by providing a formal representation of the rules of use that underwrite the natural language’s pattern of use.20

If we add a binary sentential connective governed by the tonk rules to any full language with at least one theorem and a transitive deducibility relation then the result is triviality, i.e., every theory formulable in the language will be trivial: the theory will prove every sentence in the language. Call a full language meeting this condition a trivial language and call the rules that, when added to a non-trivial language, result in a trivial language, trivial rules (N.B., a set of rules can be trivial in one context but non-trivial in another).21

---

20. If we were to define full languages more explicitly, they would be pairs consisting of languages in the standard sense together with proof theories. The terminology of a “system” is sometimes used in the literature for something like what I’m calling a “full language”: my terminology is meant to avoid confusions between languages, theories, and proof systems (theories can be assimilated to proof systems by including rules that allow us to conclude each axiom in the theory from no premises).
21. The possibility of contexts where the addition of the tonk rules doesn’t lead to triviality was first noted by Belnap (1962); it is explored in detail in Cook (2005).
Trivial rules like those for tonk seem to show that unrestricted inferentialism is ridiculous. The MRC guarantees that the tonk rules determine a meaning for the tonk connective and the MVC guarantees that the tonk rules are valid. Taken together, this results in the validity of the inference from some arbitrary sentence \( p \) to some arbitrary sentence \( q \): so, according to unrestricted inferentialism, from the sentence “2 + 2 = 4” we can infer “The New England Patriots won Super Bowl XLII”, but as every New Englander knows and laments, the Patriots, in fact, did not win Super Bowl XLII (alas). So the tonk rules are invalid and unrestricted inferentialism is false: either the MRC or the MVC must be rejected. This basic argument has convinced every philosopher who has discussed the matter, but responses have varied.

One response, seemingly favored by Prior himself, is that the MVC is to be rejected. This response allows us to hold onto the plausible idea that meanings and concepts are cheap and that the tonk rules implicitly define a meaningful tonk connective, but without the MVC, we have no reason to conclude that the tonk rules are valid and every reason to reject their validity. Above I defined “inferentialism” as the combination of the MIC and the MVC, so this response to tonk involves rejecting inferentialism. Another response, pioneered by Nuel Belnap, is to maintain inferentialism in the face of tonk by holding onto the MVC while rejecting the MRC, i.e., to retain inferentialism while rejecting unrestricted inferentialism.\(^{22}\) Let’s call proponents of this strategy restricted inferentialists.

Restricted inferentialists don’t think that the tonk rules implicitly define a meaningful tonk connective, so the MVC doesn’t force them to accept the validity of the tonk rules. If this response isn’t just to be an ad hoc attempt to avoid the oh-so rare in philosophy outright refutation, then the restricted inferentialist must say something principled about when a collection of rules can be used to implicitly define a meaningful connective, i.e., they must offer some restricted version of the MRC. The first attempt at this came from Belnap: he claimed that for rules to determine a meaning in a full language \( L^+ \) extending a full language \( L \), the condition of conservativeness must be met:

\[
\text{Conservativeness: } \quad \text{A full language } L^+ \text{ conservatively extends a full language } L \text{ if and only if } L^+ \text{ extends } L \text{ and for any sentence } \phi \text{ in the language of } L, \\
\text{if } \vdash_{L^+} \phi \text{ then } \vdash_L \phi. \quad 23
\]

As long as we start with a reasonable base (full) language, the tonk rules will fail the conservativeness constraint and so, according to Belnap, fail to determine a meaningful connective. Other post-tonk inferentialists have offered different conditions restricting the MRC, but all agree that some tonk-barring restriction is needed. This means that restricted inferentialists break from unrestricted inferentialists in denying that our rules of inference are completely self-justifying — for restricted inferentialists, there are collections of inference rules for an expression that we could follow, but that would fail to be meaning constituting. I will say more about some varieties of restricted inferentialism below, but for now the details of the various accounts aren’t of any concern.

Neither type of response to tonk is ideal: rejecting inferentialism involves rejecting a simple and plausible explanation of validity; and moving to restricted inferentialism involves rejecting the plausible idea that meanings and concepts are cheap. Moreover, the constraints offered by restricted inferentialists often make it difficult or impossible to extend the inferentialist picture from logic to domains like mathematics (see the discussion of this in section 7). Happily, unrestricted inferentialism can be maintained in the face of tonk and other trivial rules. The remainder of this paper is devoted to defending this claim. For the sake of definiteness, I will focus on tonk, but my concern here is general: I think that any collection of rules of use can be meaning constituting and that all meaning constituting rules are automatically valid.

To put things irreverently, my strategy for defending unrestricted inferentialism will involve talking with Tonkers. I’ll start with the simplest but most canonical case before moving on to more realistic tonk languages.

---

22. See Belnap (1962).

23. Belnap also added a uniqueness condition, but he ruled out tonk by way of conservativeness. As usual, \( \Sigma \vdash_L \phi \) means that \( \phi \) can be proved from \( \Sigma \) using the proof rules of \( L \).
3. Naïve Tonkers

Let’s assume, as a global inferentialist does, that natural languages like English can be modeled by full languages.\(^{24}\) For present purposes, I’m going to assume that English only includes inference rules, narrowly conceived, i.e., I’ll be assuming that all of the rules of English are standard inference rules like those in a familiar natural deduction system (this will be relaxed in section 6 below). Let \(\text{Tonklish}\) be the language that results from adding “tonk” with its normal grammar and inference rules to English. Let \(\text{Naïve Tonkers}\) be speakers of Tonklish who apply the Tonklish rules without discrimination, i.e., for Naïve Tonkers, a proof of \(p\) that uses the tonk rules is just as good as one that doesn’t and vice-versa. Tonklish is obviously a trivial language in the above sense. Is the unrestricted inferentialist’s view of Tonklish coherent?

I’ll frame the question as asking whether or not Tonklish, so described, is a possible language. By this I am not asking whether the social practice of making sounds and marks in a way that conforms to the Tonklish rules is possible; obviously, it is. I’m asking whether or not Tonklish can be understood as a meaningful language in which inferences involving the tonk rules are valid. Unrestricted inferentialists are committed to Tonklish being a meaningful language (by the MRC) and also to the tonk rules being valid in Tonklish (by the MVC); together these two points are thought to be absurd, but considering the case more carefully through an inferentialist lens will illustrate that the absurdity is merely apparent. First I’m going to discuss two immediate objections to the idea that Tonklish is a meaningful language; since these objections are intimately related to each other and fail for the same reason, I present each of them in full before replying to them in tandem. After answering them, I discuss several other objections to the possibility of Tonklish.

**Objection 1 : Truth.** The tonk rules are not valid. In Tonklish we can prove “\(2 + 2 = 5\)” (an \textit{a priori} necessary falsehood). As a matter of semantic fact, the tonk rules are not truth preserving, and so according to the just introduced terminology, Tonklish is not a possible language.

**Objection 2 : Epistemic Powers.** Relatedly, if Tonklish were a possible language then the Naïve Tonkers would have epistemic powers that they simply cannot have. Let’s flesh out this objection: assume that inference rules transmit justification so that if \(\Delta\) is a set of sentences and \(\phi\) is a sentence, and your language includes an inference rule that allows you to infer \(\phi\) from \(\Delta\), then if you are justified in accepting the sentences in \(\Delta\) and you infer \(\phi\) from \(\Delta\), you are justified in accepting \(\phi\). There are issues about the exact conditions under which inference transmits justification, but here we need only assume transmission for standard cases of inference. Let’s also assume, as a degenerate case, that if our rules allow us to infer \(\phi\) from the empty set, and we have so inferred, then we are justified in accepting \(\phi\).

Now in English we can prove (can’t we?) “either it is raining or it is not raining”; so they can prove it in Tonklish using the same route.\(^{25}\) Moreover we can prove this sentence from the empty set of premises and thus so can the Tonkers. By the above, there are situations in which the Tonkers are justified in accepting “either it is raining or it is not raining”; but now the Naïve Tonkers, using tonk introduction followed by tonk elimination, can go on to prove “the number of stars in the universe is even”. But they aren’t entitled to \textit{that}.\(^{26}\) It may or may not be true for all we know, but the Tonkers don’t have any more access to the astronomical facts than we do. It’s insane to think that one can become justified in accepting astronomical claims simply by moving from English to Tonklish. Worse still, Tonkers can prove (\textit{a priori}, necessarily) false sentences like “\(2 + 2 \neq 4\)”. In fact, because Tonklish is a trivial language, the Tonkers are in a position to accept \textit{any} sentence in their language with justification. If this were not bad enough, it seems that a plausible story of how we come to \textit{know} basic logical truths like “either it is raining or it is not raining” will carry over to \textit{every} sentence of Tonklish, so that Tonkers can know both that “\(2 + 2 = 4\)” and that “\(2 + 2 \neq 4\)”. But now the absurdity is manifest: one

\(^{24}\) For this to be at all plausible, we would need to expand the notion of an inference rule even further so as to include, e.g., acceptance of a sentence when the world is in a particular state; see Sellars (1953) for this kind of idea; in addition, recall footnote 17’s discussion of Carnap (1943).

\(^{25}\) I’m assuming that standard English is modeled by classical logic, though nothing much hangs on this assumption.

\(^{26}\) I don’t mean to mark any substantive distinction between justification and entitlement here; I’m using the terms synonymously.
can’t know something that isn’t true. We seem to face a choice: either we reject the eminently plausible and practically indispensable epistemic sanctity of inferential practices, or we reject the possibility of Tonklish.

Response: The Translation Mistake. Both of these objections rest upon a simple mistake, viz., the mistake of assuming that we can translate homophonically from Tonklish into English. A homophonic translation function \( t \) from language \( L \) to language \( K \) takes each sentence “\( p \)” in \( L \) to the syntactically identical sentence “\( p \)” in \( K \). So a homophonic translation of Tonklish into English will translate sentences in the tonk-free fragment of Tonklish directly into their English counterparts, e.g., the sentence “\( 2 + 2 = 5 \)” in Tonklish is translated into the sentence “\( 2 + 2 = 5 \)” in English. The assumption that we can translate homophonically does all of the work here, for, since the sentence “\( 2 + 2 = 5 \)” is false in English, we simply conclude that the syntactically identical sentence is false in Tonklish and so the tonk rules are invalid. But the assumption that we can translate homophonically in this case is not just false, it is obviously and horrifically false. If your favored theory of translation delivers the result that Tonklish sentences are to be translated into English homophones, then your favored theory of translation has just been reduced to absurdity.

This mistake is so important and fundamental that I’m going to give it a name:

The Translation Mistake: the mistake of being misled by superficial features into thinking that a homophonic translation is appropriate when it is not.

The unrestricted inferentialist is clearly committed to rejecting the homophonic translation from Tonklish into English, but this can be argued for on independent grounds. Any plausible theory of translation must obey some kind of minimal charity constraint. This is just to say that if our proposed translation makes those we are interpreting unaccountably wrong and shockingly irrational then our proposed translation should be rejected. How exactly to formulate the charity constraint is a difficult issue in the theory of interpretation.

Some early formulations (make the majority of the utterances of those we are interpreting true, make those we are interpreting agree with us in most cases, etc.) were unduly strong and implausible; presumably the correct form of charity will enjoin us to make the errors of those we are interpreting explicable. But whatever the details of the principle of charity, the key for my purposes is that no plausible maxim of translation will sanction the translation of Tonklish homophonically into English. Such a translation counts Naïve Tonkers as willing to utter falsehoods indiscriminately without offering a plausible explanation of their stupidity.

Note that nothing I’ve said here requires that Tonklish can be translated into English. Presumably, if Tonklish can be translated into English in a satisfactory manner, then Tonklish is a possible language. All that I’ve said here is that Tonklish can’t be translated into English homophonically and this is compatible with either Tonklish being translatable into English in a non-homophonic fashion, or Tonklish being untranslatable into English (either because of English’s defects or because Tonklish isn’t a possible language). This point gives rise to a third objection to the possibility of Tonklish.

Objection 3: We can’t translate Tonklish. We can’t translate Tonklish into English homophonically, but that’s because we can’t translate Tonklish into English at all. This obviously doesn’t establish that Tonklish isn’t a possible language, but it does shift the burden of proof onto those claiming that Tonklish is possible. If we can’t translate Tonklish into English, we should assume that our failure is because Tonklish is just meaningless gibberish: following the

---

27. For the original uses of charity principles see Quine (1960) and Wilson (1959); see Lewis (1974) and the essays in Davidson (1984) for further discussion.

28. See Dennett (1987) and Hirsch (2011) for versions of charity of roughly this kind.

29. Of course, we can only translate a fragment of Tonklish homophonically into English, namely, the fragment containing no occurrences of “tonk”.

30. Some might worry about the combination of a charity-based theory of translation with metasemantic inferentialism, since these two ideas come out of different traditions in the philosophy of language and mind. But the version of charity that I have opted for is compatible with inferentialism. In fact, use-based metasemantics, including versions of global inferentialism, go hand-in-hand with charity in translation. This is because both ideas are rooted in the insight that the meanings of sentences and expressions in the language used by a community \( C \) cannot be severed from the linguistic activities of members of \( C \). The connection between charity and use-theories has been recognized before, witness Horwich on page 72 of his (1998): “...once its precise content is elaborated, Davidson’s Principle of Charity arguably boils down to the use theory of meaning.”
Tonklish rules is obviously possible, but such a practice would not count as a “language” in any interesting sense.

Response: Yes, we can. Happily, we can translate Tonklish into English in a charitable manner. Every competent Tonklish speaker will be willing to accept the sentence “2 + 2 ≠ 4” in every situation and their willingness won’t be based on any kind of empirical test or procedure but instead on pure reasoning. The Naïve Tonkers prove “2 + 2 ≠ 4” much in the manner that we prove any simple logical or a priori truth; and what goes for “2 + 2 ≠ 4” goes for every other sentence of Tonklish as well. This suggests a strategy for securing a charitable translation of Tonklish into English, viz., translate every single sentence of Tonklish into a logical truth of English. There are several different ways to go about this, e.g., we could generate a list of the logical truths expressible in English together with a list (e.g., in alphabetical order) of the Tonklish sentences and translate the nth sentence in the Tonklish list as the nth sentence in the English list. Or we could translate each and every sentence of Tonklish using a single logical truth of English. I’ll discuss the relative merits of these various translations in the following section, but for now the differences between them aren’t important. What is important is that translating all Tonklish sentences into English logical truths provides a charitable translation of Tonklish into English.

This kind of translation doesn’t attribute any errors to the Tonkers, a fortiori, it doesn’t attribute any inexplicable errors to the Tonkers, ergo the demands of charity are satisfied. And the translation is also satisfying in other respects: in Tonklish every single sentence can be proved by pure reasoning from the empty set of assumptions; this is another way of saying that Tonklish is a language in which every single declarative sentence expresses an a priori necessary truth. In effect, every sentence in Tonklish is a logical truth.

It’s also worth noting that while we can translate Tonklish into English, English cannot be translated into Tonklish. In English we can, in a sense, express possibilities that the Naive Tonkers cannot; this despite Tonklish containing every rule that English contains and more. Tonklish is a clear case of subtraction by addition. To illustrate this, let’s think of interpretations as assigning to each sentence in a language a set of possible worlds: a sentence p is true if the actual world is a member of the set of worlds assigned to p by the interpretation (of course I’m ignoring context-sensitivity, the non-truth-apt fragments of discourse, and many other things when using this simple model). Let us say that one language has more possibility expressing power than another language if it can express some set of worlds that the other language cannot while also being able to express every set of worlds expressible in the other language. Every Tonklish sentence is assigned the set of all possible worlds, but the same doesn’t go for English; clearly English has more possibility expressing power than Tonklish.

Despite its complexity, Tonklish is an expressively weak language; for that reason, Tonklish is a useless language. But to say that a language is useless is not to say that it is impossible. In this section I have argued that were Tonklish possible, no catastrophe would result. I’ve also claimed that we can charitably translate Tonklish into English. The first point heads off several potential defeaters for the prima facie possibility of Tonklish, while the second point buttresses the case for Tonklish’s possibility. There is no metaphysical bar to speaking a trivial language however practically pointless such an endeavor would be. The next section considers further objections to the possibility of Tonklish and my translation of Tonklish into English.

4. Uniqueness, Meaning, & Translating “tonk”

This section discusses the possibility of Tonklish in more detail by answering some objections to the previous section’s discussion, focusing mainly on my proposed translation of Tonklish into English.

Objection 1: Uniqueness. The above suggested translations of Tonklish into English aren’t unique. In fact, this is an understatement: it simply doesn’t matter which logical truth(s) we map Tonklish sentences to. This “translation” is simply too undisciplined; and this lack of discipline is evidence that Tonklish isn’t meaningful.

Response. While it is true that it doesn’t matter which particular logical truth or truths we map Tonklish sentences to as far as charity is concerned (as-

---

31. That we can do this follows from the fact that both lists can be generated by algorithmic processes.
assuming all logical truths are provable, i.e., that our logic is semantically complete), this doesn’t mean that there are no constraints, e.g., any translation that maps a Tonklish sentence to a contingent sentence or a falsehood of English fails on grounds of charity. So while there is some slack in our translation, it is manifestly not the case that any translation would work. And this is evidence that Tonklish is a possible, meaningful language. In general, I don’t see the failure of uniqueness in translation as a problem. In fact, I see it as expected and innocuous. Uniqueness is too much to hope for even in real cases of translation between living, natural languages. Obviously, if any translation were acceptable, there would be reason for concern, but that isn’t the case here. The slack that exists is best seen as a reflection of the amorphousness of Tonklish vis-à-vis English.

**Objection 2: Translating Tonk.** If Tonklish is meaningful then what does “tonk” mean? The translations that you’ve suggested provide no way of translating the tonk connective itself from Tonklish into English. So what does tonk mean?

**Response.** This objection is pointing to the fact that my translations don’t translate “tonk” using a single, simple, sub-sentential connective of English. But rather than being surprising or a defect of my account, this is simply the familiar and unsurprising phenomenon of contextual definitions applied to translations. In a contextual definition sentences involving the defined expression are analyzed into sentences without the defined expression. A contextual translation of a sub-sentential expression is, analogously, a translation of every sentence involving the expression as opposed to a direct translation of the expression into a sub-sentential expression. It is hardly an objection to the account proposed here that we can only give a contextual translation of “tonk” into English. Presumably no English connective means the exact same thing as “tonk”, for if one did then English would be a trivial language.

We could, of course, give some contextual translation of sentences involving tonk that took account of the occurrence of the tonk connective, e.g., if \( t \) is our translation function taking Tonklish sentences into English sentences, we could set:

\[
t(\langle \phi \text{ tonk } \psi \rangle) = \langle (t(\phi)) \lor \neg t(\phi) \rangle \lor \langle (t(\psi)) \lor \neg t(\psi) \rangle
\]

And we could also set up our translation to be sensitive to the logical form of Tonklish sentences in general, not just those with “tonk” as main connective. We could do this, but we shan’t, for it isn’t worth doing and we end up with a contextual translation in any case. A contextual translation is enough and it is all that we should expect or hope for in cases like this. Our ability to translate Tonklish sentences including those containing “tonk” suffices for us to claim that the trivial tonk rules meaningfully define a Tonklish connective (see also objections 5 and 6 below).

**Objection 3: Compositionality.** A compositional semantics is a semantics that shows how the meanings of whole sentences are determined from the meanings of their parts. You have given no compositional semantics for Tonklish and a compositional semantics is, as many theorists have argued, a requirement on any meaningful language.\(^{32}\)

**Response.** A compositional semantics for Tonklish can easily be given. Assume for simplicity that Tonklish is a standard first-order language:

1. For any singular term \( \alpha \) in Tonklish, \( \alpha \) refers to Ø (the empty set)
2. For any \( n \)-place predicate \( R \) in Tonklish, the extension of \( R \) is \( \{<Ø,...,Ø>\} \)
3. For any \( n \)-place connective \( \oplus \) in Tonklish, \( \oplus \) is assigned a function taking all \( n \)-tuples of truth values to \( \text{True} \)
4. Quantifier expressions are functions taking appropriate items to the truth value \( \text{True} \)

\(^{32}\) Despite the popularity of the idea, I don’t think compositionality is required for meaningfulness. Sometimes the demand for compositionality is framed as a psychological claim — we can only understand new sentences in a language by grasping a compositional semantics for the language — but this is implausible: as long as the system of rules of use that give rise to a language can be implemented recursively, i.e., in some algorithmic manner, then language users could at least know how to use sentences that they had never encountered and it’s hard to see what it would mean to say that someone knows how to use a sentence perfectly in every situation without understanding it.
When truth for atomic sentences is given standardly and our domain includes the empty set, the above provides a compositional semantics for Tonklish in which every single sentence is true and its truth can be shown to be a function of the sub-sentential expressions that compose the sentence. Every atomic sentence will be true according to this semantics, and by the above clauses, all complex sentences will trivially inherit this truth value.

There is no objection to Tonklish based on failures of compositionality because Tonklish does not fail any compositionality constraint, however plausible or implausible such a constraint may be. Some may object that this semantics is somehow problematic or too easy, but I think it’s difficult to make this charge out in a cogent manner. It’s true that the above semantics isn’t unique, but what semantic theory could be unique even for a language like English? And while it’s also true that there is more slack in formulating a semantics when dealing with Tonklish, this slack is explained by the amorphousness of Tonklish and is to be expected. For these reasons, one shouldn’t try to look to this semantics as giving “the” meaning of Tonklish expressions, fixed once and for all (also for this reason odd features of this semantics can’t be used to launch objections to Tonklish’s meaningfulness).

**Objection 4:** Relative Interpretability. Whether or not two natural languages can be translated into each other is an interpretive metasemantic matter, but whether or not two full languages can be translated into each other is a mathematical matter. Logicians have long treated matters of translation and equivalence formally. The standard logical method of comparing theories is called relative interpretability and though it was used informally long before, e.g. in the study on non-Euclidean geometry, the formal treatment is due to Tarski. In the abstract, a theory $T_1$ in a language $L_1$ is relatively interpretable in a theory $T_2$ in a language $L_2$ just in case there is a translation $t : L_1 \rightarrow L_2$ meeting the following conditions,

1. All axioms of $T_1$ map to theorems (or axioms) of $T_2$.
2. $t$ preserves logical structure, i.e., $t(\phi \land \psi) = t(\phi) \land t(\psi), t(\neg \phi) = \neg t(\phi), t(\exists x \mu(x)) = \exists x (t(\mu(x)) \land t(\varphi(x))),$ etc.

When these two conditions are met not only will it be the case that $T_1 \vdash \phi$ if and only if $T_2 \vdash t(\phi)$, but also, since the logical structure of derivations is preserved, a proof of an inconsistency in $T_1$ will give rise to an inconsistency in $T_2$ (in this way interpretability results give rise to the hierarchy of consistency strength in mathematics). Typically, two mutually relatively interpretable theories are taken to be equivalent.

But the translation you offered above doesn’t meet the conditions for being a relative interpretation in the standard sense, for it doesn’t preserve logical structure. In fact, no theory formulable in Tonklish is interpretable in any consistent theory formulable in English (where we are modeling these natural languages as full languages). To see this, assume that $Tk$ is a theory in Tonklish and $E$ is a theory in English and that $Tk$ is interpretable in $E$. Then, since $Tk \vdash_{tonk} p \land \neg p$, it must be the case that $E \vdash_{Eng} t(p \land \neg p)$, so $E \vdash_{Eng} t(p) \land t(\neg p)$, and finally $E \vdash_{Eng} t(p) \land \neg t(p)$, i.e., $E$ is inconsistent: no Tonklish theory is interpretable in any consistent English theory. So no theory in Tonklish is equivalent to any consistent theory in English.

**Response.** It is true that relative interpretability in the above sense is a standard mathematical/logical method of comparing theories, but in many cases that are of interest to philosophers it’s natural to think that the conditions on interpretability should be relaxed. This is because in many cases that are of interest to philosophers we are comparing distinct logics, i.e., the two full languages will differ not just in their axioms but also in the rules the logical connectives obey or in the logical resources available in each language. What really seems to matter is that derivability be preserved and that there be some systematic relationship between the proofs under translation. My proposed translations can satisfy versions of this weaker (and admittedly vague) constraint, e.g., the kinds of translations I’ve sketched can be set up so as to preserve length of proofs, in a sense, as follows: let $n$ be the length of the shortest English proof of a logical theorem, then in Tonklish every sentence will have a proof of length $n + 2$ (some may also have shorter proofs). If “$p$” is the logical

---

33. Care must be taken in equating this formal notion of equivalence with full cognitive equivalence, for there are results showing that, e.g., proper extensions of Robinson arithmetic are interpretable in Robinson arithmetic.

34. See Tarski, et al. (1953) for the formal introduction of relative interpretability.

35. See the appendix to my (2014b).
Theorem in English with proof length \( n \), let “\( p \lor p \lor p \)” be translation of every sentence in Tonklish except \( p \) (and any other sentence having an English proof of length \( n \) or \( n + 1 \), these will map to either “\( p \)” or “\( p \lor p \)”\). This disjunction will have a proof of length \( n + 2 \), so that this translation will preserve length of shortest derivations, so we can set things up so that there is a systematic relationship of some kind between proofs in English and proofs in Tonklish.

The translations that I’ve given here fall short of some reduction or even an interpretation of Tonklish into English, but they certainly aren’t nothing. What matters in the real cases that we are intending to model is that truth conditions be preserved at least in some course-grained sense, and the translation offered here guarantees that. In any possible world, the Tonklish sentence and its English translation will have the same truth value. In addition, both sentences will be a priori and even both provable from the empty set by a proof of the very same length (if we go with the above translation). So just because the translations offered fail some important logical criteria for theoretical equivalence doesn’t mean they are philosophically unimportant or don’t serve the purposes to which they’ve been put here.

**Objection 5**: Falsity. You’ve claimed that every sentence in Tonklish is true, but absurdity is staring you in the face, for falsity is generally analyzed as truth of negation, i.e., \( \phi \) is false just in case “\( \neg \phi \)” is true. So, in Tonklish, since every sentence is true, every negated sentence is true, and so every sentence is false, ergo every sentence is both true and false, and this absurd. No sentence can be both true and false and even if some sentences can, not every sentence can.\(^{36}\)

**Response.** This would be a reductio of my position if it were right, but the objection makes the translation mistake, this time applied sub-sententially. The objection assumes that the Tonklish symbol “\( \neg \)” is, like the syntactically identical English symbol (or, rather, loglish — a mixture of English and logical symbols), a negation symbol. But this is false: the Tonklish symbol “\( \neg \)” is not a negation symbol despite obeying all of the same rules. In Tonklish, every sentence is true and no sentence is false. In order to flesh this out and make it more plausible, I need to say something about what makes some symbol in some language a negation symbol.

Some expression in some full language is a negation symbol (or, better, a negation-like symbol) just in case it plays a sufficiently similar inferential role to our negation symbol. In this way, a symbol in a Martian language, “\( \neg \)” obeying the intuitionistic rules for negation, could still be recognized as a negation symbol even though it has a slightly different inferential role in the Martian language than our symbol “\( \neg \)” has in (classical) English. The two symbols as used in their respective languages are semantic counterparts.\(^{37}\) When does some alien symbol’s inferential role count as sufficiently similar to negation’s inferential role to count as a semantic counterpart of our negation symbol? I don’t think that this question has a context-independent, interest-independent answer. Similarity is always vague, context-dependent, and is more often than not slightly indeterminate. For some purposes, intuitionistic negation and paraconsistent negation might both be considered negation-like (semantic counterparts of our negation symbol), for other purposes, we might consider only fully classical negation as negation-like.

But despite the lack of completely determinate, context-independent answers here, some symbols have inferential roles that are so dissimilar that they can never count as semantic counterparts of each other, e.g., there is no reasonable sense in which the English word “chair” could be the semantic counterpart of a Martian negation symbol. This is also the case for the English symbol “\( \neg \)” and the Tonklish symbol “\( \neg \)”\: in Tonklish this symbol has such an undisciplined use that there is no sense in which it plays the negation-role. But it isn’t that some other Tonklish symbol plays the negation role, it’s that there is simply no semantic counterpart to negation in Tonklish. How can this be the case when the Tonklish symbol “\( \neg \)” obeys all of the inference rules that our negation symbol obeys? That is the subject of the next objection; my response to it will help make sense of and justify my response to the current objection.

---

36. Dialetheists claim that some sentences are both true and false (see Priest [1987]) but no dialetheist accepts that every sentence is both true and false. In fact, Putnam (1978) has claimed that this (not every sentence is both true and false) is, perhaps, the lone a priori truth.

37. For a fuller and more general treatment of the semantic counterpart relation, see section 4 of my (2014c).
Before proceeding though, some might object that any language must include negation. I simply don’t find this plausible. It is easy enough to imagine simple languages that serve various purposes, including communicative purposes, that lack negation. Another concern might be that Tonklish can’t serve any communicative purpose, but this isn’t clear to me. A Tonker’s utterance of “p” could communicate that “p” is derivable using the rules of Tonklish (despite Tonkers being unable to formulate these thoughts in their language itself).

On some views of propositions — for example, Stalnaker’s — every logical truth in our language expresses the same proposition, the one necessary proposition (often formalized as the set of all possible worlds). The communicative purpose of uttering a logical truth, on these views, is very much like the just-given fictional story of the communicative purpose of Tonklish utterances.

**Objection 6: Incompatibility with Inferentialism.** A central tenet of logical inferentialism is that the introduction and elimination rules (sometimes just the introduction rules) for connectives like “∧” and “¬” determine the meanings of these connectives, and since Tonklish also contains the very same introduction and elimination rules for “∧” and “¬” and every other non-tonk expression in the language, this means that by the inferentialist’s own lights these expressions have the same meaning in both Tonklish and English and we should translate them homophonically from Tonklish into English. And this, obviously, guts the translation argument for the possibility of Tonklish.

**Response.** While it is true that Tonklish contains all of the introduction and elimination rules for conjunction and negation (for example) that English contains, Tonklish introduces “conjunctions” not just with conjunction introduction, but also with tonk elimination, and that makes all the difference.

This doesn’t mean abandoning the inferentialist picture, rather we maintain that the meaning of any expression in any language is given not just by the rules of use that explicitly involve the given expression, but by the total package of rules whose instances can contain occurrences of the expression even if those rules don’t explicitly involve the expression. That is: we can maintain unrestricted inferentialism if we endorse a holistic version of inferentialism.

This holism doesn’t need to be so extreme as to demand that every rule in the language for an expression is meaning constituting for the expression. In section 1 I distinguished between direct and indirect rules and claimed that it is open to inferentialists to hold that only the direct rules are meaning constituting.

The kind of holism we need to endorse doesn’t conflict with this. The idea now is only that all of the direct rules for the language are meaning constituting for the expressions in the language that can occur in their instances. And the direct rules for tonk, for example, have instances that involve the other expressions in the language, such as conjunction.

This is not an ad hoc addition to the inferentialist picture; this type of holism dovetails nicely with inferentialism’s roots in use-based metasemantics. When the tonk rules are added to a language and used as the Naïve Tonkers use them, the overall pattern of usage for all other expressions in the language changes in a significant way. If meaning is determined by use, then since the use of conjunction expressions is altered significantly by the introduction of the tonk rules, the tonk rules themselves are partially meaning constituting for conjunction and negation and other expressions. The formal point of note here is that Tonklish is not separable. Where a full language L is separable if whenever Δ is a set of sentences in the language and ϕ is a sentence in the language and Δ ⊢ ϕ, then there is an L-proof of ϕ from Δ using only the structural rules, and the rules for connectives appearing in sentences in Δ or in ϕ. To see that Tonklish isn’t separable, note that "Tonklish p ∧ ¬p" where “p” is an atomic sentence, but, since the rules of classical logic are consistent, there is no proof of “p ∧ ¬p” using only the conjunction and negation rules. And since Tonklish’s conjunction and negation rules are simply the English conjunction and negation rules and these rules are — by assumption — the classical rules, there is no Tonklish proof of “p ∧ ¬p” using only structural rules along with the conjunction and negation rules, i.e., Tonklish is not separable. What is more: all

---

39. As will become apparent in the last several sections, I am more sympathetic to objections like this than it might seem.
40. Cf. Burgess (2005)’s discussion of Dummett’s program. Even in English we introduce conjunctions without using conjunction introduction (e.g. with conditional elimination) and we even introduce the conjunction symbol into proofs without using conjunction introduction (e.g. with disjunction introduction). Burgess’s discussion looks skeptically at the possibility of justifying these practices when the view that only standard introduction rules give the meanings of connectives is taken seriously.
41. Terms like “conjunction” are ambiguous between a semantic reading and a syntactic reading; I will allow context to disambiguate in the discussion below.
Tonklish proofs of contradictions will use the tonk rules.

Is a separable language an impossible language? Many restricted inferentialists have thought so. Many of the proposed constraints on the MRC by restricted inferentialists (various types of harmony, Prawitz’s inversion principle, various types of sub-formula properties, etc.) are failed by classical logic but passed by intuitionistic logic. This has led to a general association of logical inferentialism with intuitionism, but this is by no means forced, even for restricted inferentialists. Several recent versions of restricted inferentialism, often endorsing some kind of harmony constraint on the MRC (to put things in my terms) are quite permissive.

In contrast to the restricted inferentialist, the unrestricted inferentialist doesn’t impose any constraints of this kind on meaning constituting rules, and nor should they. For as we’ve seen, conjunctions can be appropriately used in Tonklish in situations in which they cannot be appropriately used in English. This means conjunctions don’t mean in Tonklish what they do in English. The reason this doesn’t conflict with inferentialism is that some instances of the Tonklish rules contain conjunction. The meaning of an expression in any language is, for the unrestricted inferentialist, determined by the direct rules of the language that involve the expression either explicitly or in some but not all instances. If this is “holism”, then it is a holism barely worth the name. It says only that sometimes rules of use that explicitly involve some particular expression interact with rules of use that don’t explicitly involve that expression in ways that impact what can be established using the expression to such an extent that its meaning must be understood as being determined not just by the rules that explicitly involve the expression, but partly by other rules as well. There is no collapse into a “all rules are meaning constituting” picture.

Understanding the response to this objection is crucial for understanding why the homophonic translation from Tonklish into English fails. Without a grasp of this point, much of the discussion in sections 3 and 4 could seem ad hoc, which it certainly is not. This approach to meaning is essentially forced by the close tie between meaning and language use accepted by the unrestricted inferentialist. But even if all of this is accepted, we have thus far been working with very simple toy models. For recall that we’ve been considering only Naive Tonkers — people who apply the tonk rules in an utterly indiscriminate fashion and have only standard natural deduction rules in their language. What happens when we consider discriminating Tonkers who shun tonk involving proofs? And what happens when we complicate the picture by adding rejection rules? These additions change the picture in different ways: the next section deals with discriminating Tonkers and the subsequent section deals with rejecting Tonkers.

5. Sophisticated Tonkers
Above I’ve considered only Naive Tonkers — people who apply the tonk rules without discrimination. The Naive Tonkers are a sorry lot; their practice, even though it can be countenanced as a language, isn’t a very good language for any imaginable purpose. Tonklish, as spoken by the Naive Tonkers, is a language that can express no contingent information about the world.

In classical logic, the threat of inconsistency is the threat of triviality. Since I deal with triviality with such insouciance, one may wonder whether I can say anything plausible about the threat of a hidden inconsistency in our own language. On the inferentialist picture the question of inconsistency comes down to this: are the rules that govern our language inconsistent? The Church-Turing theorem tells us that there is, in principle, no algorithm for answering this question. This means that unless humans have mysterious non-algorithmic mental abilities, we cannot effectively determine whether or not the rules of our language are inconsistent (of course, it also isn’t transparent to us which rules are our rules). We can, of course, conclusively show that our rules are inconsistent by discovering an inconsistency, but without such a discovery we seem to be

42. It’s worth noting that standard natural deduction formulations of classical logic are non-separable — instances of Peirce’s Law will require non-conditional rules for their proof.
43. For a far-reaching discussion of the varieties of harmony, see Dummett (1991); see also Prawitz (1965) and (1971) for some of the most influential restricted inferentialist proposals.
44. For example, see the bilateralist inferentialism of Rumfitt (2000), the harmony proposals of Weir (1986) and Read (2000), and some of the approaches in Garson (2013).
45. More carefully: the Church-Turing theorem tells us this if Church’s thesis is true.
left with only inductive evidence that there isn’t a hidden inconsistency lurking in our rules. And, given that our evidence is finite and the number of proofs in a given formal system is infinite, whatever inductive grounds we have might be thought paltry reason for thinking our language consistent.

Once the threat of actual inconsistency and triviality has been raised, my treatment of Naïve Tonklish begins to seem worrisome. Imagine that a hidden inconsistency is lurking in our language; in such a case am I forced into saying that, contrary to appearances, we can express no non-trivial information about the world in English? It seems to be an open epistemic possibility that the rules of English are inconsistent. If the rules of English are inconsistent, then English, like Tonklish, can express no non-trivial information about the world. Since opacity isn’t an issue here, from these points it seems to follow that my treatment of Tonklish forces me into admitting that it is an open epistemic possibility that English sentences express no non-trivial information about the world.46 If correct, this objection is devastating.

I think that we can see why this objection isn’t correct by considering a different type of Tonker: Sophisticated Tonkers. The Sophisticated Tonkers, supposedly, speak a language whose rules are exactly like English plus the addition of “tonk” with its normal rules and grammar. Thus far they differ not at all from the Naïve Tonkers; the difference is this: the Sophisticated Tonkers employ their language in a discriminating fashion. While they occasionally use tonk in religious ceremonies and on other scattered occasions, in general, they don’t employ the tonk rules and they always reject tonk-involving proofs in science, mathematics, philosophy, and everyday life, i.e., the Sophisticated Tonkers don’t use the full power of the tonk rules. Let’s call the language spoken by the Sophisticated Tonkers, Sponklish, ignoring, for the moment, that Sponklish and Tonklish are rulewise-identical according to my descriptions. I am going to go on to argue that my description of Sponklish is incoherent, i.e., Sponklish doesn’t actually include the full and unrestricted tonk rules. For the moment though, let’s continue to operate under the assumption that it does.


Talking with Tonkers

Is Sponklish a trivial language? Clearly not. Sponklish seems to be just as good as English when it comes to expressing worldly information. There are tacit prohibitions against the use of certain proofs in Sponklish, and with these prohibitions in place, Sponklish is practically identical to English. How would we translate Sponklish into English? To translate Sponklish into English we should simply translate homophonically and ignore the tonk-involving fragment of the language. We should and would treat the tonk-fragment of Sponklish as a self-contained fragment that doesn’t bleed out into the larger language (this could also be accomplished, e.g., by translating all sentences in the religious, tonk-involving fragment of Sponklish, into logical truths of English).

In Sponklish, “tonk” simply isn’t a real part of the descriptive portion of the language. This isn’t a technical claim; I mean only that the Sponkers don’t use tonk in making ordinary, everyday claims about the world.

At this point an objector might point out the strain that is being put on global inferentialism here. For it seems that I’ve described two languages, Tonklish and Sponklish, with identical rules of correct use, but, because of other factors, syntactically identical sentences in the languages mean different things. If this is right, then global inferentialism is false and there is more to meaning determination than rules of use (i.e., the MIC is false). But instead of rejecting inferentialism, we should respond to this case by rejecting the idea that I’ve given a coherent description of the language and activities of the Sophisticated Tonkers. That is: despite my description, Sponklish does not include the full and unrestricted tonk rules. Only a ridiculous version of inferentialism would allow us to coherently say, for any way of filling out α, that the rules of use of community C’s language are R and yet the members of C act α-wise. Some ways of acting, e.g., some patterns of verbal behavior, must be incompatible with speaking a language constituted by rules R.

This is most perspicuously represented using the notions introduced in sections 1 and 2: my presentation of Sponklish involved a stipulated pattern of use U and a stipulated full language L+, but the rules of use of a language (represented here by the proof theory of L+) were supposed to explain the pattern of use of the language, U. But in this case, clearly the rules of L+ don’t explain the pattern of use U. This is particularly apparent given the account of rule-following endorsed in section 1, but on any reasonable view of linguistic
rule-following, realizing some patterns of use will be incompatible with following certain rules. So the Sponkers do not speak a language that includes the full tonk rules. The problem with the description of Sponklish I gave above is that I tried to stipulate that Sponklish involves certain rules while also stipulating that Sponkers act in a way incompatible with following those rules. My description was simply incoherent. Sponklish isn’t Tonklish, Sponklish is (nearly enough) English, i.e., in saying that Sponklish includes the full and unrestricted tonk rules, I said something false.

I’ve argued that we can translate Sophisticated Tonkers homophonically and Naïve Tonkers non-homophonically, but there are cases that don’t fit neatly into either of these boxes. For a representative example, consider Forgetful Tonkers — speakers who are disposed to follow the tonk rules, without these dispositions ever being realized. Unlike the Sophisticated Tonkers, it seems that Forgetful Tonkers really do follow the tonk rules; but unlike the Naïve Tonkers, it seems appropriate to translate them homophonically.

How we should treat the Forgetful Tonkers depends on further details of the case. If Forgetful Tonkers really and truly would accept a proof of "P = NP", using the tonk rules, if it were presented to them, then they are simply Naïve Tonkers. Their disposition to exploit the tonk rules, even if unrealized, is enough to undermine homophonic translation of their language into ours. At first glance, this might seem a bit odd, but consider the following case: your friend is interested in the existence of extraterrestrial life. In fact, he is so concerned that he spends most of his time volunteering for SETI and using his home radio to listen for non-natural radio signals from space. Imagine now that the Oracle tells you that, even though your friend has never considered the matter, he would, clearheadedly, accept an argument exploiting the tonk rules to the conclusion “extraterrestrial life exists” to be just as good as hearing radio signals from Martians or some other form of proof. I think it’s clear that, in this case, you would be forced to conclude that, despite appearances, your friend’s comments about “extraterrestrial life” don’t mean exactly what they seem to mean. The Forgetful Tonkers are like this but a million times over. They’re disposed to accept a tonk proof for anything, and so they don’t mean what we mean.

If this still seems weird, it’s because we are tacitly imaging the Forgetful Tonkers as something like Sophisticated Tonkers. Imagine we stipulate that if you were to remind the Forgetful Tonkers about a tonk-involving proof of "P = NP", they would — seemingly — admit that the proof was valid, but wouldn’t treat the proof as a mathematical breakthrough (no Fields medal for you). In such a case, despite superficial appearances, the Forgetful Tonkers are just Sophisticated Tonkers because the tonk rules aren’t really a part of their language. So, depending on how we flesh out the case, the Forgetful Tonkers are either Naïve or Sophisticated. Either way, Forgetful Tonkers (and similar variant cases) don’t pose unsolvable problems for unrestricted inferentialism.

Consideration of these cases naturally leads us from Tonkers to ourselves. Here is an argument that the rules of English are inconsistent: consider the Liar Paradox. The Liar is a sentence \( L \) that says of itself that it isn’t true, i.e., let corners here be a naming device and let \( T \) be the truth predicate, then:

\[
L : \sim T L
\]

It is highly plausible that our grasp of the concept of truth is (at least partly) constituted by the following rules:

\[
\text{Truth Introduction : } \frac{\phi}{T \phi} \quad \text{Truth Elimination : } \frac{T \phi \sim}{\phi}
\]

I think nearly every English speaker would endorse these rules in their unrestricted form; but these rules, together with the Liar and classical logic, lead to the result that English is inconsistent and therefore trivial. Either \( L \) or \( \sim L \), but if \( L \) then both \( \sim T L \) by definition and \( T L \) by truth introduction; and if \( \sim L \) then \( T \sim L \) by definition and double negation elimination and then also \( \sim T \sim L \) by truth elimination and definition. Thus, English is inconsistent.

When presented with the Liar Paradox, ordinary speakers politely smile.

---

47. Another example is provided by Introduction Tonkers — speakers who often use tonk’s introduction rule but never use tonk’s elimination rule.
JARED WARREN

Talking with Tonkers

and shrug in mild indifference. They think that the Liar is a fun puzzle of little importance. They think so, presumably, because as English speakers we tacitly adopt prohibitions against Liar proofs much as the Sophisticated Tonkers tacitly prohibit tonk proofs. If I were to tell a group of astronomers that I had discovered that the number of stars in the universe is even, at best, they would politely ask to see my evidence. If I then used the Liar and the explosiveness of contradictions to prove my claim I would soon be escorted out by security and beaten severely. Nor would I ease my plight by challenging them to point out exactly where my reasoning went awry. What this illustrates is that a hidden inconsistency would only threaten our language with triviality if we were disposed to exploit the inconsistency, once discovered, as a runabout inference-ticket. Since we are not so disposed, the hidden threat of inconsistency is a chimera.

In the case of the Sophisticated Tonkers, cordoning off the unproblematic fragment of the language is quite easy. By contrast, in the case of English’s truth rules, there isn’t an obvious way to reframe the rules so that they don’t lead to trouble save for simply saying that our truth rules hold for any sentence in the language except for liar-like sentences. Things get tricky here, since there are technical problems with constructing theories of truth simply by screening out problem cases. It is an open question whether a satisfactory theory of truth in English could be constructed along the lines indicated, but all that is important for my purposes here is the philosophical point that because English speakers are not disposed to exploit contradictions as runabout inference-tickets, there is no real epistemic possibility that English could be a trivial language like Tonklish.

The case of the Sophisticated Tonkers illustrates that with a proper view of rules and use, there is no epistemic possibility of English turning out to express no non-trivial information about the world. This is perfectly compatible with my treatment of Tonklish. Once an inferentialist perspective is adopted, the question of the consistency and non-triviality of our rules becomes pressing. The question becomes less pressing once we realize that since there is no real possibility of our being disposed to accept explosive proofs and thus triviality, there is no real danger of our language turning out to be trivial. Our plight is not the plight of the Naïve Tonker, it is the plight of the Sophisticated Tonker.

6. Rejection Tonkers

So far I’ve been working with a model of natural languages according to which all rules are standard single-conclusion rules of inference. Normatively, these can be seen as rules of acceptance: they sanction accepting the conclusion when you accept all of the premises. In general, standard inference rules tell us how to expand the number of sentences we accept. Since standard natural deduction systems include only rules of this kind, it isn’t like I’ve made a crippling or ad hoc assumption in focusing on inference rules in this narrow sense. But as was noted in section 1, while inference rules in this narrow sense are important rules of language use, it’s plausible that they aren’t the entire story.

For instance, it’s plausible that actual natural languages are sometimes modeled by using in addition to acceptance rules, rules of rejection. These rules might tell us to reject some sentence $\phi$ in such-and-such circumstances. To employ a metaphor: the mental state of accepting a sentence $\phi$ is roughly that of having the content of the sentence in your belief box (ignoring context sensitivity and other issues here), but the mental state of rejecting a sentence $\phi$ isn’t simply not having the sentence’s content in your belief box: when you suspend judgment as to $\phi$, you don’t have the content of $\phi$ in your belief box, but you aren’t rejecting it either. Rejection is stronger than mere suspension of judgment; it involves actively excluding a sentence’s content from your belief box.

What happens when we add rejection rules to the picture? Let’s assume we’re working with a more fine-grained model of English that we’ll call “R-English” (Rejection English), and assume further that we move from R-English to R-Tonklish (Rejection Tonklish), where, as before, R-Tonklish simply adds the tonk rules to R-English. Let’s assume also that there is some sentence $p$ (“Rudolf Carnap is a fried egg”) that the rules of R-English enjoin us to reject: is R-Tonklish a possible and meaningful language?

48. The main result I have in mind here is from McGee (1992).

49. Formulations of natural deduction systems employing rejection as well as acceptance are offered in Smiley (1996) and Rumfitt (2000).
I don’t think so for the following reason: the R-Tonklish rules will tell us to both reject \( p \) (because the rules of R-English tell us to reject \( p \)) and also to accept \( p \) (because the R-English rules tell us to accept, e.g., \( \lnot p \lor \lnot \lnot p \) and “\( p \)” is provable from this sentence in R-Tonklish). But whatever else it is to accept \( p \), it is to be disposed to act in certain ways (e.g., employing the content of “\( p \)”, in that guise, in your deliberations, etc.); and whatever else it is to reject \( p \), it is to be disposed to act in certain ways (e.g., not employing the content of “\( p \)”, in that guise, in your deliberations, etc.). And these ways will directly conflict with one another, so it isn’t possible to both accept and reject the very same thing. This is relatively uncontroversial: even dialetheists like Graham Priest, who think that some sentences can be both true and false, don’t think that it is possible to both accept and reject the very same sentence. 50 Acceptance and rejection are mental states that you can only be in if you have certain patterned behavioral dispositions, but you can’t instantiate the acceptance pattern for \( p \) at the same time you instantiate the rejection pattern for \( p \), so you can’t be in both the mental state of accepting \( p \) and the mental state of rejecting \( p \) at the same time. So it isn’t possible to follow the rules of R-Tonklish.

I think that this point can be used to show that R-Tonklish isn’t a possible and meaningful language, at least if we accept the plausible principle that “ought implies can”:

**Ought Implies Can (OIC):** for any agent \( S \), for any action \( X \): if \( S \) ought to \( X \), then it is possible (in principle) for \( S \) to \( X \)

With OIC together with the above discussion of R-Tonklish and acceptance and rejection, we can give a simple argument for the impossibility of R-Tonklish, assuming \( p \) is salient:

1. If R-Tonklish is possible, then for some sentence \( p \), the R-Tonkers ought to both accept and reject \( p \) (by the definitions of R-English and R-Tonklish)
2. For any sentence \( p \), it is not possible for R-Tonkers (or anyone else) to both accept and reject \( p \) (argued for above)
3. So: R-Tonklish is not possible (1,2, OIC)

50. See Priest (2006), page 103.

N.B., in distinction to the above discussion of standard Tonklish, here we are rejecting the possibility of R-Tonklish not because we are rejecting unrestricted inferentialism’s neat and tidy naturalistic metasemantic picture by rejecting either the MRC or the MVC, but rather because it simply isn’t possible to follow the R-Tonklish rules. The impossibility isn’t owing to some external constraint in metasemantics, rather, it’s rooted in our capacities as physical beings in a physical world. Using the terminology introduced above, the problem with R-Tonklish is that the formal object of the full language involving both rejection rules sanctioning the rejection of \( p \) and the tonk rules, corresponds to no possible pattern of use given uncontroversial facts about normativity and the mental states of acceptance and rejection.

We’ve seen that both Sponklish (under my initial description) and R-Tonklish are not possible languages, but their impossibility has nothing to do with the falsity of unrestricted inferentialism. They are impossible because the descriptions of the languages show that no agents could, in principle, be correctly described as following the stipulated rules of use while instantiating the pattern of use that they are said to instantiate. In the Sponklish case this is because the stipulated pattern of use, while possible to realize, isn’t generated by the supposed rules of Sponklish; in the R-Tonklish case it is because the supposed rules of R-Tonklish cannot, as a matter of principle, generate any pattern of use. The upshot of the last two sections is that a naturalistic, dispositionalist account of rule-following combines with inferentialism to avoid truly absurd results, e.g., that English might express no contingent information about the world or that it is possible to both reject and accept a sentence. Now let’s take a step back and discuss unrestricted inferentialism vis-à-vis its competitors.

7. The Merits of Unrestricted Inferentialism

I’ve argued that Tonklish is a meaningful language and that the tonk rules implicitly define a meaningful connective. I’ve also argued that this position is perfectly compatible with inferentialism and a sensible view of actual inconsistency. Here I will briefly point to two potential advantages that unrestricted inferentialism has over its metasemantic competitors: it (i) promises to lead to a simple, naturalist metasemantics and (ii) opens the door to applications to
mathematics and other branches of discourse.\footnote{There are other issues facing unrestricted inferentialism that I haven’t discussed here, e.g., there is a kind of converse to the tonk-problem. But the issues of triviality and inconsistency are generally and, in my opinion, rightly considered to be the most significant obstacles to unrestricted inferentialism.}

(i) Naturalist Metasemantics. Despite the differences in rhetoric, I think my position here will be congenial to many restricted inferentialists. In fact, some may object that there isn’t much of a real difference between the two versions of inferentialism. This objection claims that we are simply talking about different possible social practices and disagreeing about whether they deserve the honorifics “language” or “meaningful.”

When considering some versions of restricted inferentialism this objection has merit, but many versions of restricted inferentialism suggest a very different picture of how semantic properties fit into the natural world. Unrestricted inferentialists start with syntactic inferential practices involving the deployment of whole sentences and then, on that basis, go on to explain sentential semantic properties like truth. With both syntactic inferential practices and sentential semantic properties in hand, the unrestricted inferentialist goes on to explain sub-sentential semantic properties like reference.\footnote{For an attempt to carry out this step for reference in mathematics, see Wright (1983).} Naturalist inferentialists would then go on to explain our syntactic and inferential practices in purely scientific terms, e.g., in terms of complex patterns of dispositions and causal interactions; anti-naturalist inferentialists would explain our syntactic and inferential practices using one or more non-semantic but non-naturalized primitives, e.g., the anti-naturalist may appeal to primitive normative notions.\footnote{Brandom (1994) seems to accept a kind of anti-naturalist inferentialism involving primitive normativity.} But both naturalist and anti-naturalist versions of unrestricted inferentialism ultimately explain semantic notions by appeal to our inferential practices.

The picture suggested by some versions of restricted inferentialism, by contrast, sees the failure of the tonk rules as analogous to reference failure. The expression “The female who was President of the United States in 1950” fails to refer to anyone, since the description isn’t satisfied by anyone. According to this picture, the tonk rules act much like a definite description — they specify which sub-sentential meaning is picked out, if any, by the tonk connective. If there is no meaning in reality to make the tonk rules truth-preserving, then the term “tonk” experiences something akin to reference failure.\footnote{But this picture makes our inferential practices beholden to semantic notions and thus reverses the order of explanation endorsed by unrestricted inferentialists.}

The non-inferentialist response to tonk (accept the MRC while rejecting the MVC) is very similar to the restricted inferentialist response just sketched: the non-inferentialist sees the tonk rules as picking out a meaning for “tonk” but thinks that this meaning doesn’t validate the tonk rules. Once again, metasemantics is not autonomous. We can pick out meanings by following rules but it is up to the world to determine which of our meaning-constituting rules are valid. I’m not claiming that restricted inferentialism and non-inferentialism cannot be made compatible with a naturalistic view of semantics, but it’s clear that unrestricted inferentialism (in its naturalistic form such as I have endorsed here) is — at first glance — a much better fit than either of its competitors with the standard scientific picture of the world.

(ii) Mathematical Conventionalism reborn?\footnote{The very idea of truth by convention faces many challenges that I don’t discuss here; for defense from the most influential objection, see my (2014a).} Unrestricted inferentialism also has appealing potential applications that restricted inferentialism and non-inferentialism don’t share. For just one example: the constraints on the MRC imposed by restricted inferentialists, e.g., that our rules lead to conservative extensions of our current theories, often make it difficult or impossible to apply the inferentialist story in domains like mathematics. Using Belnap’s version of restricted inferentialism that was discussed in section 2: any formulation of number theory will entail the counting sentence of first-order logic “there are at least \(n\) things” for every \(n\), and unless our base theory already entails that infinitely many things exist, this means that our arithmetical rules will non-conservatively extend our base theory, and so can’t be meaning constituting. But if unrestricted inferentialism is endorsed, we can take collections of rules that don’t conservatively extend our base theory to be meaning constituting and thus automatically valid for our arithmetical expressions.
To illustrate: we could give a rule-based formulation of the standard Peano axioms for number theory. Imagine, to start, that we’re working in a standard first-order language and we add to it a monadic predicate “N” (is a number), a constant “0" (zero), and a one-place function s(x) (“successor of”) with the following inference rules governing their deployment:

(P1) \[ \alpha \vdash N\alpha \]
(P2) \[ \frac{\alpha \vdash N\alpha \quad \alpha \vdash \neg s(\alpha)}{\alpha \vdash \neg \exists x \ Ns(x)} \]
(P3) \[ \frac{\alpha \vdash \exists x \ Ns(x)}{\alpha \vdash \exists x \ Ns(x)} \]
(P4) \[ \frac{\alpha \vdash \exists x \ Ns(x) 
 \alpha \vdash s(\alpha) = s(\beta) 
 \alpha \vdash \alpha = \beta}{\alpha \vdash \beta} \]
(P5) \[ \frac{\phi(0) \quad \forall \xi (\phi(\xi) \rightarrow \phi(s(\xi))) 
 \alpha \vdash \phi(\alpha) \vdash \alpha \vdash \phi(\beta)}{\alpha \vdash \phi(\beta)} \]

Let’s call these the Peano rules — in a standard natural deduction system for first-order logic these rules will generate each of the Peano axioms (with the proviso that the induction “axiom” is really an axiom schema). Conversely, if we start with the Peano axioms in a standard natural deduction system we can prove each of the above as derived rules. And if we add to the above system the standard clauses for the addition (+) and multiplication (\(\times\)) functions, we get full, standard first-order Peano arithmetic.\(^{56}\)

The unrestricted inferentialist, but not restricted inferentialists, can claim, by the MRC, that the Peano rules are meaning constituting rules for our arithmetical vocabulary (the number predicate, the zero constant, and the successor function). And the unrestricted inferentialist, but not the non-inferentialist, can claim, by the MVC, that the Peano rules are automatically valid (necessarily truth-preserving). This allows unrestricted inferentialists to use these rules to explain the truth of any arithmetical sentence that follows from these rules, e.g.,

the following is a proof that two is a number, according to the Peano rules (the proof uses P1 followed by two applications of P2):

\[ \frac{\alpha \vdash \forall \xi (\phi(\xi) \rightarrow \phi(s(\xi))) \quad \alpha \vdash \phi(\alpha) \vdash \alpha \vdash \phi(\beta) 
 \alpha \vdash \phi(\beta) \]

Since, by the MVC, all of these rules are necessarily truth-preserving, the sentence “\(Ns(s(0))\)” is true. In this way, the rules of our language alone explain why arithmetical sentences are true (at least for provable sentences). Nowhere in the above explanation of the truth of “\(Ns(s(0))\)” did we need to appeal to facts about the world that went beyond the rules of our language. In short: accepting unrestricted inferentialism re-opens the possibility of extending inferentialism to conventionalism about mathematics.\(^{57}\) This is a potentially fruitful application of metasemantic ideas that is unique to unrestricted inferentialism.\(^{58}\)

In saying this, I am not claiming that all versions of restricted inferentialism will block the application of inferentialism to arithmetic and other mathematical subjects. Some less restrictive constraints on the MRC may allow for arithmetical inferentialism or even versions of set theoretic inferentialism. But there are two further points to make in favor of unrestricted inferentialism against any type of permissive restricted inferentialism. The first is that by laying down restrictions in advance, we may run into cases where, intuitively, inferentialism should apply, but according to our constraints, it does not. And it would, for example, be very odd if we could apply inferentialism to some set theories that are in wide use but not others. We would be in the position of either revising our constraints on an entirely \textit{ad hoc} basis or being put into the uncomfortable position, as philosophers, of telling working mathematicians that they’ve started to talk nonsense. This isn’t merely an idle worry, for the

\(^{56}\) These axioms are (i) \(\forall x(x + 0 = x)\), (ii) \(\forall x \forall y (x + s(y) = s(x + y))\), (iii) \(\forall y (y \times 0 = 0)\), and (iv) \(\forall y \forall z (y \times (x + z) = y \times z + y)\). They could each be added in rule form to be more in keeping with inferentialism.

\(^{57}\) The account just sketched would only explain the truth of provable sentences of arithmetic. The unrestricted inferentialist would need to supply some account of Gödel-style true but unprovable sentences; see Gödel (1931).

\(^{58}\) For further discussion of mathematical conventionalism, see my (2014c). Also, see footnote 61 below.
Neo-Fregean program in the philosophy of mathematics has run into this type of situation in attempting to extend the Neo-Fregean approach from arithmetic to set theory.\textsuperscript{59}

The second point is that, as noted briefly in section 1, even permissive versions of restricted inferentialism have dropped the central idea that the inference rules we use are completely self-justifying. For example, harmony constraints are often based on the idea that introduction rules are self-justifying, but the analogous claim for elimination rules is dropped entirely. The unrestricted inferentialist — but not the permissive restricted inferentialist — maintains the original motivation for inferentialism in an entirely undiluted form: the rules of use for our language need answer to no external tribunal, they are entirely self-justifying. So even if some versions of restricted inferentialism can apply inferentialism to arithmetic, no version of arithmetical conventionalism would be forthcoming. This is because the restricted inferentialist’s explanation of the truth of a provable arithmetical claim is not simply the rules of our language alone, but also includes that fact that our rules meet whatever external condition is imposed. By contrast, unrestricted inferentialists leave no hostage here, so the possibility of various kinds of logical and mathematical conventionalism is re-opened.\textsuperscript{60}

Of course, one philosopher’s \textit{modus ponens} is another philosopher’s \textit{modus tollens}, and some may see the possibility of conventionalism as an evil to be avoided at all costs. But if unrestricted inferentialism has non-absurd potential ramifications in the general philosophy of logic and mathematics, the theoretical attractiveness and comprehensiveness of the view is increased. Showing that conventionalist ramifications are non-absurd is a task for another occasion.\textsuperscript{61}

My main goal in this paper has been a kind of demythologizing: our concepts of language and meaning aren’t inscribed in Platonic heaven. I’m perfectly willing to admit that in some contexts and for some purposes we won’t want to treat Tonklish as a meaningful language. For some classificatory purposes the Tonkers’ Tonklish utterances will count as more similar to complicated animal noises or the roar of the sea than to any human language; but restricted inferentialists make risky transcendent rulings about which “languages” are meaningful and which social practices are “languages” and these transcendent verdicts risk being implausible. Restricted inferentialism risks turning tractable pragmatic prohibitions into mysterious metaphysical edicts. We should be able to reject the tonk rules and other problematic rules without claiming that agents involved in a social practice using such rules cannot be expressing any meaningful claims. Unrestricted inferentialism, but not restricted inferentialism, allows us to do this.

8. Practical Choices

There is still a practical problem facing the global unrestricted inferentialist: the addition of the tonk rules to English causes semantic recoupling, i.e., sentences are re-interpreted in the move from English to Tonklish. Worse, they are re-interpreted in a way that makes every sentence inter-derivable, which makes the move disastrous. Clearly we should \textit{not} add the tonk rules to English. But which rules should we add? And which rules — other than the tonk rules — should we shun? This is a pressing practical problem. We are constantly introducing new expressions and altering the way that we use old expressions. Scientists consider adding new theoretical terms implicitly defined by scientific theories, mathematicians consider introducing terms for new kinds of mathematical objects implicitly defined by a set of axioms, and ordinary speakers are assaulted daily with the question of which new terms and concepts to add to their linguistic and conceptual stores.

The unrestricted inferentialist thinks that before us lies the boundless ocean of unlimited possibilities, but clearly not every possibility should be actualized. And while I think we should go along with Carnap in advocating \textit{theoretical} tolerance, in many practical situations \textit{intolerance} is reasonable. Is there anything to say about the problem of which rules we should or shouldn’t add that is

\textsuperscript{59} For Neo-Fregeanism, see Wright (1983) and Hale & Wright (2001); for the problems with bad company, see the essays in Linnebo (2009).

\textsuperscript{60} I won’t discuss this here, but I don’t think the use of logic and general metasemantic principles in explaining mathematical truth undermines conventionalism.

\textsuperscript{61} A full theory of both logical and mathematical conventionalism, built on a foundation of unrestricted inferentialism, is presented in my forthcoming book, \textit{Shadows of Syntax}.
both helpful and general? There is, I believe, a kind of master prohibition, viz.,
we tend to frown upon alterations to our language that violate the following
condition:

**Homophonic Translation**: A set of rules $R$ added to a full language $L$
(along with appropriate expressions) satisfies the *homophonic translation constraint* if and only if we can (according to our norms of translation) translate homophonically from the old full language $L$ into the new full language $L + R$.

Of course there are various dimensions along which this constraint can be partially satisfied, but there is good reason for adoption of some form of the constraint: if we cannot translate homophonically into our new language, then we will not be able to take for granted that the sentences we previously accepted should still be accepted after the change. We will, so to speak, have to start over epistemically. And starting over is hard work, so we tend to embrace expansions of our language only if the change preserves our old reasons for accepting our old sentences. The homophonic translation constraint partially and imperfectly codifies this demand; it captures the demand to the extent that it does because the type of meaning preserved under translation suffices for evidential equivalence in most if not all circumstances. But the only way to tell if the constraint is satisfied is by appealing to our best theories of translation; there is no mechanical test.

Despite its practical importance, the homophonic translation constraint isn’t *metaphysically* sacrosanct — it isn’t that alterations of our language that fail the constraint are somehow meaningless. Instead, the point is that we often have practical reasons to avoid embracing meaningful expansions of our language. What’s more: the homophonic translation constraint isn’t even *practically* sacrosanct — there are times when the practical advantages offered by a new language, say in expressive power, are worth some epistemological dovers. Decisions about which expansions to accept can’t be made in advance; they must be taken on a case-by-case basis and be sensitive to the practical roles that particular segments of discourse play in our lives.

This point must be stressed, for without it, it might be assumed that in tentatively endorsing the homophonic translation constraint, I am simply endorsing another version of restricted inferentialism. But I am not; homophonic translation is not a constraint on meaning constituting rules. It is not even an all-things-considered constraint on the practical development of our language; sometimes I think it makes sense to adopt a new rule that violates the constraint. For these reasons, the move from metaphysically important metasemantic constraints on meaningfulness to a *ceteris paribus* practical constraint is not mere idle relabeling.

### 9. Conclusion
Unrestricted inferentialists need not fear tonk. It’s open to the unrestricted inferentialist to maintain that Tonklish is a possible and meaningful language and that the tonk rules implicitly define a meaningful “tonk” connective. There is no reason for inferentialists to claim that the tonk rules metaphysically misfire. One of the most appealing features of inferentialism is its rejection of external metaphysical constraints on linguistic meaning, and this feature can be maintained even in the face of trivial rules.

Tonklish, at least as spoken by the Naïve Tonkers, is a poor language. The language is expressively impoverished relative to English and would be practically useless for most imaginable purposes. But to say that Tonklish is a bad language is not to offer some kind of metaphysical condemnation. Building a metaphysical idol to the evils of tonk accomplishes nothing save for obscuratanism. To claim that a language is practically useless is condemnation enough.\(^{62}\)

---

\(^{62}\) For comments, thanks to Paul Boghossian, David Chalmers, Cian Dorr, Hartry Field, Katrina Przyjemska, Jack Woods, Crispin Wright, an audience at NYU, and several referees. Thanks also to Daniel Waxman for LaTeX help.
References
Garson, James W. (2013). What Logics Mean: From Proof Theory to Model-


