Making It Implicit

Inferentialism and the Limits of Internalising Meaning

by Niklas Dahl

Bachelor's Thesis
Course code: FTEK01
Thesis advisors: Erik J. Olsson
Faculty examiner: Robin Stenwall

Presented on February 27, 2018.

Contents

1 Introduction

The meaning of sentences is a notoriously tricky beast to nail down. In spite of the ease at which we manage to understand new and unfamiliar expressions every day most of us would be at a loss if asked what that understanding consists in. In contemporary philosophy linguistic meaning is often conceived of as a relation of referring or representing things in either some possible world or the actual one. Another strand of thought, often pursued in opposition to the first, instead puts the spotlight on the roles that sentences can play in reason and action. This second approach is the starting point of *inferentialism* or *inferential-role semantics*. Broadly speaking this view says that the meaning of a sentence is captured by its connection to other sentences in reasoning. A particularily detailed inferentialist theory of meaning has been put forward by Robert Brandom and has since been much debated. In this text I give a birds eye view of this brand of inferentialism as well as it's alleged shortcomings. In particular I focus on an avenue of attack, originally due to Kevin Scharp, which targets the tension between the goals of explaining the meaning of semantic vocabulary and being able to state the theory of meaning for a language inside that language. After reiterating the exchange between Brandom and Scharp I offer a sharpening of the original objection. The tension between these goals has so far been discussed in terms of an inferentialist theory of semantics but through a use of Lawvere's Fixed Point theorem can be shown to extend to all attempts of internalising theories of meaning.

2 Inferentialism in a General Context

When discussing linguistic meaning we must begin by offering a distinction between two questions: 'What is the meaning of particular sentences?' and 'What systems or practices can be linguistically meaningful?'. The first question concerns what Dummett (1991, pp. 20-22) calls a "meaning-theory" for a particular language and the second what he calls the "theory of meaning". I will deviate slightly from Dummett in using 'semantic theory' rather than 'meaning-theory'.[1] In this text the focus will be on theories of meaning. Where theories of meaning concern themselves with general ideas, about how it's possible to convey information to others, semantic theories deal with the actual meaning of expressions in a particular language.

Following Dummett's (1991) discussion on theories of meaning we can make some distinctions between possible semantic theories. One way to partition the bestiary of semantics is by considering what Dummett calls "the central notion of a [semantic] theory" (1991, pp. 32). This is the notion in terms of which we can explain the validity of arguments. An obvious candidate for a central notion would be to have the truth-conditions of sentences play this role. After all, weren't we all taught in our basic logic classes that preservation of truth was precisely what characterises valid inferences? This is the species to which Davidson's (1967) semantic theory belongs. The main idea of truth-conditional semantics is to borrow from the model theory of logic a recursive definition of truth-conditions for complex sentences in terms of simpler ones. Such a semantic theory for a language $\mathcal{L}$ is adequate if it for each sentence S in $\mathcal{L}$ can generate a T-sentence for S. That is a sentence

$$\text{`}S\text{' is true in } \mathcal{L} \text{ if and only if } P.$$

where P expresses the truth-condition of S. As a result of Davidson's long and productive career the literature is full of discussions and criticisms of his semantic theory so, as it is not relevant for this text, I will move on. [2]

In contrast to the semantic theories whose central notion is truth Dummett considers theories in which the central notion is, as in the Wittgensteinian aphorism, 'use'. Since we use sentences in varied ways and roles he turns to Frege's work on language to make explicit the different ingredients of meaning as use. In his monumental exegesis of Frege's work (1981, pp. 84-86) Dummett presents three distinct parts determining how an expression can be used. The 'sense' of an expression is the part that relates to the determination of the semantic value

[1] While this use is more standard it does conflict with Dummett's use of the term. In his use a semantic theory is the assignment of semantic values to sentences modeled on semantics of formal logic. These sorts of theories make up the "base" of meaning-theories (Dummett, 1991, pp. 25).

[2] The reader interested in criticism of truth-conditional theories can turn to Soames (1992), Speaks (2006), or Dummett (1991). In fact Dummett spends a large part of the book on facts about the theory of meaning to conclude that truth-conditional theories aren't proper semantic theories.

of the sentence.[3] We can think of it as what the content of the expression would be if used assertively. The ‘force’ of an expression instead determines what kind of speech-act is being performed by producing the expression. Consider the following statements:

$$\text{Tea is a superior drink to coffee.} \tag{1}$$
$$\text{Is tea a superior drink to coffee?} \tag{2}$$

The expressions share a sense as they express the same idea. However they differ in that (1) is an assertion while (2) is a question. Producing (1) might have as an appropriate response either a simple agreement or a renouncing of your friendship if your interlocutor happens to hold particularily strong views on warm beverages. An utterance of (2) would instead have as appropriate responses different ways of giving an answer to it. It follows that we can’t entirely determine the pragmatic significance of an expression without already knowing its sense. Even if someone might understand that the expression offered is a question they can’t know how to properly respond without knowing what is asked. Dummett therefore characterises types of force as being different ways to determine the pragmatic proprieties of an expression given its sense.

The final ingredient in the meaning of expressions is what Frege calls ‘Tone’. Frege offers us the example of the distinction between ‘and’ and ‘but’ used conjunctively. They are used to convey different information but it is not a difference of speech-act or semantic value. However a clearer definition is not forthcoming. As Dummett puts it:

> It serves to define the proposed *style* of discourse, which, in turn, determines the kind of thing that may appropriately be said. We may speak to one another solemnly or light-heartedly, dispassionately or intimately, frankly or with reserve, formally or colloquially, poetically or prosaically; and all these modes represent particular forms of transaction between us. [...] When a dictionary notes, after its definition of a word, ‘archaic’, ‘vulgar’, or the like, it is, quite properly, indicating tone. (Dummett, 1991, pp. 122, emphasis in original)

Because of the difficulties in nailing it down a theory of ‘tone’ is generally passed over in silence hoping that some later philosopher will deal with the issue. In this text I will also follow this well-trodden path.

The next step in such a semantic theory is then to see what sort of thing can be the sense of an expression. With the added vocabulary that Frege gave us we can consider the truth-conditional view as a theory of sense without explicitly corresponding views on force and tone. Consequently if we consider truth-values

[3] In Freges original formulation the ‘sense’ of an expression is what determines the truth-value of a sentence since Frege assumed that truth and falsehood were the only possible semantic values of a sentence (Dummett, 1991, pp. 114).

to be the correct semantic value then by following Dummett's recipe we would just reproduce the truth-conditional theory.[4] In order to find an alternative Dummett considers what we must know or master in order to competently assert an expression (1991, pp. 83-86). In an earlier text he puts it more explicitly:

> Learning to use a statement of a given form involves, then, learning two things: the conditions under which one is justified in making the statement; and what constitutes acceptance of it, i.e., the consequences of accepting it. (Dummett, 1981, pp. 453)

To build on this idea, in the same way that the Davidsonian approach borrowed its main idea from model theory, it is possible to turn to another branch of logic: Proof theory.

Structural proof theory is an approach to logic in which the central objects of study are formal deductive systems.[5] These are collections of schematic rules that recursively define the notion of a proof. The theory then proceeds to investigate the relative strength of such systems and their combinatory properties. The idea relevant for my purposes comes from the fact that in these systems every logical constant comes equipped with two distinct kinds of rules: Introduction and elimination. I will give as examples the rules for 'and' ($\wedge$)[6]:

$$\frac{\varphi \qquad \psi}{\varphi \wedge \psi}\wedge I \qquad\qquad \frac{\varphi \wedge \psi}{\varphi}\wedge E \qquad \frac{\varphi \wedge \psi}{\psi}\wedge E$$

This means that we are allowed to conclude that 'φ and ψ' if we already know both of them and that, conversely, if we know 'φ and ψ' then we can conclude each of 'φ' and 'ψ' in turn. In the paper introducing natural deduction Gentzen (1934, pp. 80) tells us that "The introductions represent, as it were, the 'definitions' of the symbols concerned [...]". That is to say that, according to Gentzen, the entire meaning of the symbol '$\wedge$' is contained in the rule $\wedge I$.[7] It is this idea

[4] Strictly speaking this is not the case. In fact Dummett does approach the problem by assigning 'truth-values' in a sense but moves from a Tarski-like semantics which requires classical two-valued logic to discussing Kripke semantics (as a general approach to non-classical semantics and not just modal logic). The goal of this is to show that different formal semantics give rise to different valid rules of inference. The discussion of the justification of logical laws, while interesting in its own right, is not directly relevant for my purposes and so we take the shortcut to a direct characterisation of sense through inference. An interesting aside is that many of Dummett's thoughts about non-classical semantic values seem formalised by the idea of a subobject classifiers in Topos theory. After a cursory search I have found no expositions on the matter but this subject would be far to lengthy an aside to develop here.

[5] An introduction and examples can be found in Troelstra and Schwichtenberg (2000).

[6] Dummett (1991) prefers in his exposition to use the formalism of sequent calculus. This has the added benefit of making it easier to state the deductive rules of your choice, especially in regard to differing contexts of assumptions. As we will only be explicitly discussing very simple rules they will instead be stated in natural deduction for ease of reading.

[7] According to Gentzen it is always possible to derive, in some sense, the elimination rules for a connective from the introduction rule. Dummett (1991) takes this idea seriously for more general expressions and develops a theory on the justification of inferences based on Prawitz (1974). Dummett and Prawitz also both take this to be a symmetric state of affairs and that by considering the elimination rules as basic the introduction rules can be 'derived'.

that Dummett extrapolates to a proper semantic theory by, instead of defining the meaning for a particular symbol, defining the sense of an entire expression through the inferential rules that governs it. The idea that the meaning of an expression being specified through (some of) the inferential relations it stands in is the foundational idea of *inferential-role semantics* or *inferentialism*.[8] Choosing different collections of inferences to be considered meaning-constitutive gives rise to different semantic theories. Dummett illustrates the situation by offering a theory he labels as "verificationist", in which the sense of an expression is determined entirely by its introduction rules, and one he labels "pragmatist", in which the sense is instead determined entirely by elimination rules.

The idea of meaning as constituted by inferential relations originates independently[9] in Sellars' (1953) work on concepts. He considers it clear that the validity of the following argument in some way rests on the meanings of the sentences involved.

$$\frac{\text{Socrates is human}}{\text{Socrates is mortal}}$$

He discusses such arguments in relation to philosophers of logical positivism who would claim that the preceeding argument is an enthymeme. Expressed clearly the argument would rather be:

$$\frac{\text{Socrates is human} \qquad \text{All humans are mortal}}{\text{Socrates is mortal}}$$

What makes the original inference valid in terms of content is then that there was a suppressed premise, the correctness of which was guaranteed by its meaning, that made the inference valid as a matter of logic. Sellars puts their view as follows:

> Without formal rules of inference there would be no terms, no concepts, no language, no thought. In this sense, our empiricists continue, one could say that logical rules of inference specify, at least partially, the very form of a term or concept. (Sellars, 1953, pp. 314)

In this way the logical positivists wish to subsume all correct inferences into formally valid inferences. The conceptual content of a sentence P is then relegated to analytically true conditional sentences involving P. In Sellars' view this is a mistake. Instead of characterising meaning in terms of what conditional sentences are valid we can instead turn to what inferences about it are valid. Then

[8] In the larger bestiary of semantic theories inferential-role semantics is a special case of *conceptual-role semantics*, which can allow for conceptual content to be formulated in other terms than pure inferential relations, and a generalisation of *proof-theoretic semantics*, which takes logical proof to be the central type of inference. Proof-theoretic semantics is developed in detail for first-order logic in Prawitz (2006).

[9] Sellars' work does not mention, or betray awareness, of the proof theoretic approaches of Gentzen. In fact the formal logic in the relevant papers is done either in terms of a Hilbert calculus or the formalism of Principia Mathematica.

the original argument about Socrates' fate would be valid in itself independently of whether that inference can be codified in a conditional sentence. This lets him define the broader notion of *material inference*; which are those rules of inference that are valid as a result of the meanings of the concepts involved.

Sellars takes material rules of inference as essential to describe the meaning of expressions. He thinks that they can't properly be subsumed into logical inferences with suppressed premises. He argues that logical inference can't capture the counterfactual robustness of conceptually licensed inferences. First '$\forall x$ x is red $\rightarrow$ x is coloured' doesn't license that 'If x were red then x would be coloured' which is an inference we ought to consider valid. On the other hand universally quantified conditional sentences express something too strong for certain conceptually licensensed inferences. The inference from 'I struck this match' to 'This match was lighted' is one that we would endorse only with a suppressed *ceteris paribus* clause. The additional circumstance that the match was wet would block this inference. But given the formal conditional 'If a match is struck then it is lighted.' would then either be rendered false by the situation where a match was wet or license the conclusion that the match was not struck. In either case the formal expression has not captured the conceptual content of 'match' properly. Since Sellars finds counterfactual talk to be an essential part of our language it would follow that not all proper inferences could be reduced to classical conditional statements.[10].

Whether or not we agree with Sellars that material rules of inference are essential to our semantic vocabulary he has at the very least shown us that it is possible to adopt an inferential perspective on the role of conceptual content in licensing inferences. With the vocabulary of materially valid inferential relations we have a wider conception of valid inference to consider as meaning-constitutive for a semantic theory. With this broad overview of how inferentialist semantic theories can be conceived we turn to a particular brand of inferentialism worked out in detail by Robert Brandom.

3 Brandom's Theory

Before going into detail on Brandom's theory of meaning there are two of his methodological commitments to discuss:

(1) The point of a semantic theory is to explain the correct use of linguistic expressions. (Brandom, 1998, pp. 133)

(2) The meaning of complete sentences has explanatory priority. (Brandom, 2007, pp. 651-654)

The first of these points is similar to the one made by Dummett above that to know what an expression means requires mastering it's use. From this Bran-

[10]A much more detailed discussion of this argument and the meaning of counterfactuals can be found in Sellars (1957).

dom argues that grasping the meaning of an expression requires grasping the inferential relations it stands in:

> The parrot does not treat "That's red" as incompatible with "That's green", nor as following from "That's scarlet" and entailing "That's colored". [...] What the parrot [...] lack[s] is an appreciation of the significance their response has as a reason for making further claims and acquiring further beliefs, its role in justifying some further attitudes and performances and rule out others. (Brandom, 1998, pp. 89)

These sort of performances, using expressions as reasons and justifications, are ones that Brandom takes as basic to our use of propositionally meaningful language (1998, pp. 158). On this view the ability to exclaim a sentence only when it is appropriate is not sufficient to understand its meaning. To further explain the correct use of linguistic expressions it is however not sufficient to know all the inferential relations between sentences as that would at most give a theory of 'sense' in the Fregean terminology. To fully explain 'use' Brandom needs a theory of force. To this end Brandom explicitly invokes Dummett's conception of, types of, force as uniform patterns of deriving practical significance from the meaning of the expression (Brandom, 1998, pp. 189). Therefore explaining 'sense' must be the first building block in constructing a semantic theory.

The point (2) also has its origin in a commitment to explaining semantic meaning in terms of some aspects of its use. The meaning of entire sentences must come first in the explanatory project simply because complete sentences are the minimal units of language which we can use to make moves in a language-game[11] (1998, pp. 79-82). Taking entire sentences as the explanatorily basic parts of language also fulfills another function for Brandom. Although I will not discuss it at length here, this move allows Brandom to ground his semantics in a theory of linguistic practices.[12] For the purposes of this text I will take as primitive the proprieties of material inferences.

In light of these preliminaries I can offer a strategic overview of the theory. The first step is to define meaning, as 'sense', for complete sentences. From there Brandom uses substitution to distinguish kinds of subsententials, like singular terms and predicates, and consider their sense through what difference their occurence makes for the complete sentences they occur in. With the additional resources from that part it is possible to explain the meaning of more general expressions containing anaphoric tokens. Finally after having explained the sense of general expressions a theory of force can be added to the mix according to Dummett's recipe.

[11] Brandom discusses the speech-acts of referring or naming as possible counterexamples. However he thinks that without having first explained the meaning of entire expressions, containing the name or designation, this doesn't give any meaning to the name or designation as it would not give us access to any new performances.

[12] For a detailed explanation of Brandom's theory of linguistic practices the reader is referred to the first chapter of *Making It Explicit* (1998).

3.1 Constructing Semantic Theory

The initial step is the following definition for the meaning, as 'sense' in Frege's terminology, of a complete sentence:

Definition 1. The *meaning of a sentence* consists of the, broadly construed, inferential relations it occurs in. (Brandom, 1998, pp. 131)

The first thing to note with is that Brandom, unlike Dummett, does not pick out either the circumstances of application (I-rules) nor the consequences of its application (E-rules) as the foundational sort of inferences (Brandom, 1998, pp. 121-123). The reason for this is that he finds both kinds of inferences to be necessary to specify the content of a sentence. Consider again the above example of the parrot who reliably can report the presence of red objects. As stated above it can't be considered to have grasped the meaning of its exclamations on the redness of things. Consider, conversely, a person who knows all the consequences of someone being considered a criminal but, through a very particular reading of legal texts, has no idea whom to properly apply that term to. Such a person could not be considered to in general grasp sentences containing the word 'criminal'.

The next step to understanding this definition is to clarify what Brandom means by inferential relations 'broadly construed'. First he adopts Sellars' notion of material inferences as the foundational sort of valid inferences. That 'Pittsburgh is to the west of Princeton' follows from 'Princeton is to the east of Pittsburgh' is the sort of inference that is meaning-constitutive for both the involved sentences (Brandom, 2000, pp. 52-53). However material inference as construed by Sellars does not, in Brandom's view, suffice to deal with the different attitudes that we can have with respect to the assertion of a sentence (Brandom, 1998, pp. 159-161). Given a particular assertion we can evaluate our commitment to it as well as our entitlement to it. This allows us to make a distinction between material inferences which preserve commitment[13], as the one from 'Princeton is to the east of Pittsburgh' to 'Pittsburgh is to the west of Princeton', and those which preserve entitlement, such as the one from 'I struck this match' to 'This match was lighted'.[14] What makes entitlement-preserving inferences special is that their conclusions are defeasible.

To characterise the final sort of material inference Brandom considers the relation of incompatibility between assertions.

Definition 2. Two assertions P_1, P_2 are *incompatible* if commitment to one precludes entitlement to the other. (Brandom, 1998, pp. 160)

[13]Considering logical entailment as a kind of commitment-preserving inference has the added benefit of giving us the ability to explain why we might think that if $P \to Q$ and we know P then we ought to know Q. Clearly we aren't always aware of all logical consequences of our beliefs but we are committed to them (Brandom, 2000, pp. 174).

[14]These species of material inferences are not disjoint. Any inference preserving commitment must also preserve entitlement (Brandom, 1998, pp. 191). Otherwise we could be both committed and entitled to some assertion P but be only committed and not entitled to some consequence of P.

The incompatibility relation is what bridges the gap between commitment and entitlement. Imagine that we are both committed and entitled to the assertion 'This match was struck' as well as committed to the assertion 'This match is wet'. Even though our entitlement to 'This match was lighted' follows from the first assertion it is blocked by its incompatibility with the second assertion to which we are committed. Incompatibility also allows us to pick out another kind of material inference.

Definition 3. The assertion P_1 *incompatibility-entails* the assertion P_2 if every assertion incompatible with P_2 is incompatible with P_1. (Brandom, 1998, pp. 160)

This is the notion of entailment that Brandom takes as basic.[15] It can be considered a straightforward generalisation of standard logical consequence where incompatibility between P_1, P_2 takes the role of $P_1, P_2 \vdash \bot$.[16]

The second part in explaining the "broad construal" of inference is to allow for the possibility of empirical content of sentences. This comes from the fact that we are sometimes entitled to some assertions not because of some other assertion but as a result of reporting what we perceive. In the presence of a cup of tea I am entitled by default to assert 'There is a cup of tea in the room' but not to assert 'There is a rhinoceros in the room'[17]. This can be considered broadly inferential as my entitlement to such assertions is the conclusion of an inference whose premises are my sensory perceptions (Brandom, 1998, pp. 224-225). In this way the disposition to reliably infer entitlement to a sentence from the non-inferential circumstances reported by our senses in the appropriate situations is part of its meaning. The point is to think of it as inferring entitlement to P from the fact, as observationally reported, that P. These sort of assertions then base their default entitlement on the fact that the asserter understands the meaning of the sentence[18] (Brandom, 1998, pp 227).

[15] This is where Brandom's normative theory of pragmatics is to do the explanatory work of determining what material inferences and incompatibility relations are correct.

[16] Write $P \vdash_I Q$ for the relation that for all sentences R if $Q, R \vdash \bot$ then $P, R \vdash \bot$. Let R be any sentence such that $Q, R \vdash \bot$. If $P \vdash Q$ then $P, R \vdash \bot$ so $P \vdash_I Q$. On the other hand if $P \vdash_I Q$ then in particular $Q, \neg Q \vdash \bot$ and so by hypothesis $P, \neg Q \vdash \bot$. Hence $P \vdash Q$, assuming a classical logic. The first half of the argument generalises seamlessly to Brandom's broader concept of incompatibility-entailment and so giving a 'soundness' theorem for intuitionistic (and classical) first-order logic over Brandom's system.

[17] Assuming, pace Wittgenstein, that there is no rhinoceros in the room. Distinguishing which observational assertions we are by default entitled to becomes for Brandom a matter of the pragmatics of challenging and vindicating entitlement. My default entitlement to some claims is a matter of my reliability to make such claims only when proper. While I might be expected to reliably distinguish situations containing tea from ones which, catastrophically, do not there is no-one who would, fortunately, treat me as reliable in distinguishing what ails a particular patient by asking me for medical advice. Treating me as a reliable claimer of P can also be given an inferential formulation. It means treating the inference from my assertion P as entitling your own commitment to P (Brandom, 1998, pp. 220-223). The pragmatics of reliability are treated at length in Chapter 4 of *Making It Explicit* (1998).

[18] Brandom illustrates this point by holding that the entitlement to the assertion 'There is a cup of tea in the room' can then be defended in the appropriate circumstances by, as Wittgenstein, answering "I speak English".

Coordinate with considering the meaning of some sentences containing the appropriate perceptions entitling their assertion is the idea that the meaning of some sentences contains inferences from their commitment to actions[19]. Brandom gives an example:

> In a culture in which white is the color of death, and things associated with death are to be shunned or avoided - a culture, to be sure, that would mean something somewhat different than we do by their word corresponding to our ‘white’ - the connection between the visible presence of white things and the practical response of shunning or avoiding, which their practitioners endorse by using the concept in question, is an inferential one in the broad sense in question here. (Brandom, 2007, pp. 658)

As in the case of perception, which consists of inferring entitlement from extralinguistic circumstances, Brandom considers normative attitudes to actions to be extralinguistic, practical, consequences of commitment to some assertions. Similarily to the default entitlement of observational assertions there are default commitments and entitlements to act in a certain way that follows from certain claims. For example, in the culture described in the quote, there follows a practical commitment to avoid a thing from commitment to ‘That is white’. The disposition to reliably[20] act in appropriate ways as a result of certain commitments is again part of the, broadly inferentially construed, meaning of those sentences.

With the definition of meaning made explicit Brandom offers a categorisation of the parts of meaning as *conceptual* content, contained in inferences between sentences, *empirical* content, contained in inferences from perception, and *practical* content, which is contained in the inferences whose consequences are proprieties of action (Brandom, 1998, pp 234). The next step in extending the notion of meaning is to deal with subsentential expressions.

The key to offering meaning for subsentential expressions is through what substitutional inferences they figure in. This is done by essentially moving backwards from the form of logical substitution-inferences. In the case of singular terms they have the form:[21]

$$\frac{t \doteq s \qquad \varphi[x/t]}{\varphi[x/s]}\, S$$

[19]By actions here are meant intentional acts. Brandom borrows a framework taken from Davidson (1963) that intentional actions are performances which have a description that is the conclusion of some piece of practical reasoning. In Brandom’s terminology that corresponds to there being some sentence being expressed by the action that we are committed or entitled to make-true (Brandom, 2000, pp. 93-95).

[20]In the case of perception we can only be entitled by default if we are reliable observers in the relevant way. Analoguely, we can only be considered committed to act in a particular way if we are considered reliable in producing that act. Since my piano playing skills are next to non-existent there is no claim I can be committed to which would result in me being committed to play the Moonlight Sonata as that is not an act I can reliably perform (Brandom, 1998, pp. 235-236).

[21]The notation $\varphi[x/t]$ denotes the sentence φ in which occurences of x is replaced by t.

The argument is that since the terms t, s are equal we can infer $\varphi[x/s]$ from $\varphi[x/t]$. Brandom, on the other hand, begins his account by taking some collection of inferences to already be materially valid. Then he can pick out which terms can be substituted for each other by seeing whether both of

$$\frac{\varphi[x/t]}{\varphi[x/s]} \qquad \frac{\varphi[x/s]}{\varphi[x/t]}$$

are materially valid inferences. If this is the case for all sentences where t occurs[22] then we could conclude that the meaning of the two terms is the same. As an example; consider the following inference:

$$\frac{\text{Mark Twain wrote Huckleberry Finn}}{\text{Samuel Clemens wrote Huckleberry Finn}}$$

That this inference, and the one in opposite direction, is valid is part of what the meaning of 'Mark Twain' consists. The entire meaning of 'Mark Twain' consists of all other terms which are in symmetric substitutional relations with it. Being a singular term, according to Brandom, is to be a subsentential expression whose substitution inferences are symmentrical.

Moving on to substitution of general subsentential expressions Brandom considers another class of inferences. Naturally a first condition of such substitutions is that they preserve sentencehood; no permitted substitution will transform a sentence into some ungrammatical gibberish. We call the equivalence class of intersubstitutables for some expression the *type* of that expression. An example of a valid substitution would be:

$$\frac{\text{Socrates is human}}{\text{Socrates is a featherless biped}}$$

Here what was substituted was the sentence-frame in which Socrates figured. This inference differs also from the previous ones in that it is not symmetric, as demonstrated by Diogenes of Sinope. According to Brandom this is what distinguishes predicates from other types of subsentential expressions (1998, pp. 372). The meaning of a predicate $Q(x)$[23] is then all inferences of the forms:

[22] The criterion Brandom puts forward doesn't require that s and t are intersubstitutable for precisely all sentences in which t occurs. Sentences like 'Jones believes that Mark Twain wrote Huckleberry Finn' and 'Jones believes that Samuel Clemens wrote Huckleberry Finn' are not to help us determine which singular terms have the same meaning. That this substitution inference might not be materially valid even though the terms 'Mark Twain' and 'Samuel Clemens' have the same meaning should not block us from claiming that they have. Brandom proposes that we distinguish these irrelevant inferences by the fact that knowing whether $s \doteq t$ or not doesn't determine whether these inferences are valid (Brandom, 1998, pp. 373-374).

[23] I have taken the liberty of considering predicates, as in Lewis (1970), as functions taking as input singular terms and producing complete sentences. While Brandom doesn't put it that way he does claim to show that the framework of functional-categorial grammars, as considered by Lewis, is equivalent to his approach. (Brandom, 2007, pp. 674; Brandom, 1998, pp. 404-409)

$$\frac{P(x)}{Q(x)} \qquad \frac{Q(x)}{R(x)}$$

These examples of giving meaning to subsentential expressions indicate a way to generalise the notion of meaning to arbitrary subsentential expressions. If φ is an arbitrary subsentential expression then denote by $\varphi|\psi$ the sentence yielded by putting it together with some other expression ψ, of appropriate type, in the appropriate way.

Definition 4. The *meaning of a subsentential expression* φ consists of the, broadly construed, inferential relations that sentences of the form $\varphi|\psi$ stands in with sentences $\rho|\psi$, where ρ is a subsentential expression of the same type as φ.

With the meaning of subsentential expressions in hand the next phase is Brandoms explanation of anaphora. The subsentential expressions I've discussed so far have all been repeatable in the sense that their meaning is the same for different utterances of them. This distinguishes them from anaphoric expressions like 'It' and 'Her'. While anaphoric expressions have a grammatical type, in the sense that we know what class of subsentential expressions they are intersubstitutable with, their meaning is not determined by simply considering inferences between sentences they occur in. The meaning of the word 'Here' depends crucially on where I am when I utter a sentence containing it.[24] Hence there is a need for theory to understand the meaning of such expressions in terms of the meaning of repeatable expressions.

The approach to take here is to realise that every meaningful anaphoric expression inherits its content from some other meaningful expression. In particular what is inherited is the validity of substitution inferences containing the anaphoric expression. To put it more clearly: Let ϵ be an anaphoric expression inheriting its content from an expression ψ. What this means is that all valid inferences containing the expression ψ remain valid with ϵ substituted for ψ (Brandom, 1998, pp. 454-455). As an example take the following inference:

Sappho was a lyric poet. Therefore she wrote poems.

The 'She' here stands in exactly the same substitutional relations as 'Sappho'. Assuming the same anaphoric antecedent for 'She' this makes the following substitution-inference correct:

$$\frac{\text{Sappho wrote the `Ode to Aphrodite'}}{\text{She wrote the `Ode to Aphrodite'}}$$

When we have established what is the antecedent for an anaphoric expression then we know precisely what repeatable expressions we can substitute for it. Since we have already explained the meaning of such expressions in terms of substitution this extends the notion of meaning to anaphora as well.

[24] In Brandoms discussion of anaphora he takes indexicals and deixis to be special cases of more general anaphoric structures (Brandom, 1998, pp. 458-486).

The immediate antecedent of an anaphoric expression does not generally need to be, as in the example above, non-anaphoric. If it is another anaphoric expression then it too inherits its meaning from some expression. This allows for chains of anaphoric reference. However such dynasties of inheritance must begin somewhere. Following the content of an anaphoric expression back through antecedents we at some point find an expression, or demonstration, which serves as an anaphoric initiator (Brandom, 1998, pp. 458).

3.2 Force as a Scorekeeping Practice

With Brandoms theory of sense in hand I can give a brief explanation of his conception of force. As noted above he agrees with Dummett that:

> [C]orresponding to each different kind of force will be a different uniform pattern of derivation of the use of a sentence from its sense. (Dummett, 1981, pp. 361).

The question is then what these sort of uniform patterns consist of. In order to give an answer Brandom borrows an idea from David Lewis; extending the notion of a language-game to contain a practice of scorekeeping. Lewis (1979, pp. 339-341) argues that at every point in a conversation there are statements which are presupposed by what is said. Sometimes they are explicitly asserted as when someone claims that 'Tea is truly the supreme warm beverage' and in other cases they are implicitly presupposed. The second case can be illustrated by the example that 'The poetry of Emily Dickinson was published posthumously' additionally informs us of the fact that Emily Dickinson is no longer alive. Lewis talks of these different presuppositions as an example of the *conversational score* in analogy with the score of a game.[25] The score of the conversational game tells us what presuppositions are in play and thereby what utterances can be made without further justification. That the score is updated in a rule-governed way is what allows us to communicate information implicitly as in the Dickinson example.

This general model of conversational practice is adapted by Brandom in order to give a theory of force. Harking back to the earlier discussion of material inferences the distinction was made between commitment and entitlement to a sentence. These attitudes are what Brandom takes to be possible scores in the game of conversation. When talking to each other we keep track of what statements each participant in the conversation is committed to as well as which they are entitled to. This collection of attitudes for each participant is what Brandom calls the *deontic score*[26]. Brandom agrees with Lewis that the practices for updating the deontic score needs to be rule-governed. It is precisely

[25] To make the analogy clearer it might be better to consider what Lewis is talking about as the state of a game rather than the score. While the number of points scored, at some particular time, by each team in a game of football is part of what Lewis is describing so are all other pieces of information needed to reconstruct the game at that time.

[26] Putting it into a more familiar idiom the score consists of all the beliefs and entitlements to belief of each participant in the conversation.

in these rules that Brandom (1998, pp. 188) identifies the force an utterance. A particular rule for updating the deontic score given the sense of an utterance is a particular type of force.

Definition 5. A *type of force* is a rule for updating the deontic score.

To illustrate this definition an example is in order. What makes a particular utterance S by an speaker a into an assertion is that a participant, b, in the conversation updates their scoreboard in the following way (Brandom, 1998, pp. 190-191):

(1) b attributes to a commitment to S.

(2) b attributes to a commitment to all sentences which are commitment-inferential consequences of S and other claims that b attributes to a.

(3) b ceases to attribute to a entitlement to all sentences which are incompatible with the commitments attributed in the previous steps.

(4) b evaluates whether or not a is entitled to S.[27] If b takes a to be entitled to S then that entitlement is attributed to a.

(5) If b considers a entitled to S then b attributes to a entitlement to all claims that are entitlement-inferential consequences of S together with other claims that b attributes to a unless the consequence is incompatible with any commitment that b attributes to a.

(6) If b considers a entitled to S then b attributes entitlement to S to each participant whom b does not attribute commitment to a claim incompatible with S.[28]

This rule for updating the scoreboard is what the force of an assertion consists of. Other kinds of speech-acts correspond to other rules for altering the deontic score. In the rule defining the force of assertion several of the steps required us to know which inferential relations the claim stands in. Therefore we need to know the meaning, in the sense of sense, of a claim in order to precisely determine the force of expressing it. Rules (such as the one above for assertion) on how to update the deontic score are then precisely the "uniform patterns of derivation" that Dummett mandated. This allows for a natural definition of force for particular utterances.

Definition 6. The *force of an utterance* is the difference that utterance makes to the deontic score. It is determined by its sense together with its type of force.

[27] Entitlement to a claim, according to Brandom, can come in several ways. First it can be the result of a entitlement-preserving inference from some claim that person is already entitled to. Secondly it can come about through the testimony of others entitled to it. Thirdly it can be defended as the result of a reliable disposition to report the claim in the correct circumstances. Finally it can be a claim to which we are entitled by default. An in depth discussion of entitlement can be found in *Making It Explicit* (1998, pp. 176-178, pp. 199-229)

[28] On first glance this last step may seem out-of-place. The idea is that whenever we take someone makes an assertion, which they are justified in making, that is a licence for others to assert the same claim.

3.3 The Expressive Role of Normative Vocabulary

Looking back at the inferentialist theory of semantics proposed by Brandom one might notice that, while filled to the brim with inferential relations, the theory makes no use of expressly *logical* vocabulary or inference. Instead it relies entirely on Sellars notion of material inference. Unlike theories which begin with a notion of logical vocabulary and inference as primitive Brandom begins instead with material proprieties of inference and uses them to identify the underlying logic of the language. His approach is as follows: Start by fixing a collection of logical symbols. *Logical inferences* are then those inferences which are materially valid under all substitutions of non-logical vocabulary. The equivalence classes formed under the relation of being a substitution instance of each other then tell us the rules of inference for the logic underlying the language. However in order to use this method we would need to already know what vocabulary is to be considered logical. In order to complete his account of logic Brandom needs to specify how to pick out specifically logical vocabulary.[29]

What characterises logical vocabulary, according to Brandom, is that it allows us to express the inferential structure of meaning within the language. The case of the conditional, $\rightarrow$, is the most striking example. Consider the following argument:

$$\frac{\text{Antigone was a very early proponent of civil disobedience}}{\text{Antigone was concerned with questions of virtue}}$$

Without access to the conditional this argument can only be put forward implicitly by stating the two sentences in such a way that they are, in practice, treated as inferentially linked. Adding a conditional to this practice allows the argument to be made explicit as a sentence to which others can attribute commitment or entitlement. The role of logical connectives is to allow speakers to put forward the inferential relations between sentences as judgeable contents (Brandom, 1998, pp. 112).[30] Brandom treats the other connectives in a similar way. Negation is to make explicit incompatibility relations, equality makes

[29]The line of thought used to pick out logical inferences can be expanded to pick out inferences valid in terms of their K-form. Fix a collection K of symbols and categorise the K-valid inferences as those which are materially correct under all substitutions of symbols not in K (Brandom, 1998, pp. 104-105). Similarily there is an approach, not followed by Brandom, to pick out specifically logical vocabulary as the collection K which yields as K-valid inferences those which are correct under all 'isomorphisms' of the relevant structures. This approach has recently experienced an upswing in interest as a result of the monumentous discovery of Homotopy Type Theory and the related Univalence Principle. A discussion of this approach and its history can be found in Awodey (2014).

[30]This is also precisely the role that $\rightarrow$ plays in deductive systems. The formula $\varphi \rightarrow \psi$ can be judged as true or false were the judgement $\varphi \vdash \psi$ can not. However given a correct judgement $\Gamma, \varphi \vdash \psi$ the deduction theorem tells us that this is equivalent to $\Gamma \vdash \varphi \rightarrow \psi$. Taking this equivalence as defining the conditional gives rise to precisely the rules for $\rightarrow$ in Gentzen's system of natural deduction. The other connectives can be handled in a similar way as inferences between judgements in the metalanguage. In this way proof-systems can be considered as an internalisation of the logic of the metalanguage into the object language (Troelstra & Schwichtenberg, 2000, pp. 28-31). Brandom's approach is similar except that it substitutes our implicit linguistic practice for a formal metalanguage.

judgable whether two terms are intersubstitutable, and quantifiers express inferential relations between predicates. Having access to such logical vocabulary then enables reasoning about the contents of expressions which were previously implicit in their use.

This conception of logic is available to Brandom because he treats material inference as given to us through linguistic practice. If he had instead treated material proprieties of inference as enthymemes then their validity would need to rest on a prior understanding of logic. By allowing for a more general notion of inference and then characterising logicity through expression Brandom gets to explain which inferences and vocabulary is logical rather than using them as explainers.

Having characterised logic as making explicit conceptual content Brandom moves on to more practical normative vocabulary. Consider the following piece of practical reasoning:

$$\frac{\text{By eating a biscuit I will be happier}}{\text{I shall eat a biscuit}}$$

Just as in the case of conceptual reasoning we can either treat this inference as correct on its own merits or by relying on a suppressed premise like 'I should do things which make me happier'. Above I discussed that the alternative order of explanation allows us to treat the "suppressed" as a sentence expressing the validity of certain inferences. This in turn allowed the logical vocabulary to be defined through what inferences it makes explicit. Brandom takes the same approach to normative vocabulary such as 'prefers', 'must', and 'ought'. The role of these words is to allow us to put forward sentences codifying implicit practical reasoning. To assert that 'I wish to be happier' is simply to endorse inferences as the one above.

3.4 Explaining Truth

An additional requirement that Brandom undertakes for the success of his theory is that it can explain the standard uses of representational vocabulary such as 'True' (1998, pp. 279). Since these expressions are part of everyday usage he thinks that any self-respecting semantic theory ought to be able to explain their meaning. Additionally Brandom seeks to offer a story about "objective representational content" in terms of which the correctness of claims can be asserted independently of our attitudes toward them (1998, pp. 54).

The approach to truth that Brandom endorses builds on the one proposed by Grover, Camp, and Belnap (1975) to treat truth-ascriptions as *prosentences*. Prosentences are tokens which inherit their meanings from the sentence that is its anaphoric antecedent. They are defined analogously with how pronouns inherit their meaning from some noun. The meaning of the sentence

'Snow is white' is true.

is then precisely the same as its antecedent 'Snow is white'. This type of evaluation they call *lazy* (Grover, Camp, & Belnap, 1975, pp. 84). For more complicated sentences such as

Everything Jane said is true.

the first step is to decouple the quantifier into a collection of things, in this case sentences that Jane said, and then evaluating the truth-ascription lazily on each of them. Hence the quantified sentence has the same meaning as asserting each of the sentences that Jane said (Grover, Camp, & Belnap, 1975, pp. 94-95). The final step is to deal with sentences of the form

Goldbach's conjecture is true.

In this case simply removing the truth-ascription doesn't result in an assertion. To proceed we need to rewrite the sentence in some way to remove the expression 'Goldbach's conjecture'. As it is a pronoun it is an expression which inherits its meaning from an anaphoric antecedent which in this case is another sentence. In fact there are many sentences, all of them equivalent, which might correspond to that pronoun. As such we can rewrite the sentence as

All sentences referred to by 'Goldbach's conjecture' are true.

and then deal with the resulting sentence in the quantificational way (Grover, Camp, & Belnap, 1975, pp. 95).

To this theory Brandom wishes to make a small alteration. Instead of the quantificational approach above he prefers to consider the quantificational expression as an anaphoric token picking out what sentence(s) the *prosentence-forming operator* '... is true.' is applied to. Looking at the example above 'Everything Jane said ...' is then picking out what sentences S are asserted by the original claim.[31] To summarise Brandom's view "... is true." is considered an operator that takes as input some expression naming a collection of sentences and then producing as output a sentence which anaphorically inherits it's meaning from that collection.

4 Objections to Inferential-Role Semantics

After having sketched the ideas of inferential-role semantics in general, and Brandoms brand of inferentialism in particular, I can now turn to its detractors. Objections to inferentialism come in many flavours. Some question whether meaning can be inferentially articulated at all while others note how standard semantic worries still seem to plague the inferentialist enterprise.

[31] In discussing the operator view of the theory Brandom never takes as example a universally quantified sentence and speaks of the referring expression as "picking out *the* tokening on which the whole prosentence depends" (Brandom, 1998, pp. 305, emphasis mine). This continues from this point onward which is even more strange considering that he explicitly uses universally quantified examples just a few pages earlier.

4.1 Tonk and Inferential Meaning

An early critic of meaning being entirely inferential, even in the restricted case of logical connectives, was A.N. Prior. In a paper (1960) he aquaints us with the logical connective 'tonk'. The rules for introducing and eliminating it are as follows:

$$\frac{P}{P \text{ tonk } Q}\, I_1 \qquad\qquad \frac{Q}{P \text{ tonk } Q}\, I_2$$

$$\frac{P \text{ tonk } Q}{P}\, E_1 \qquad\qquad \frac{P \text{ tonk } Q}{Q}\, E_2$$

This curious connective then has the introduction rules of 'or' and the elimination rules of 'and'. Hence in a language containing tonk the following deduction is valid:

$$\frac{\dfrac{P}{P \text{ tonk } Q}\, I_1}{Q}\, E_2$$

In such a language it is always the case that $P \vdash Q$ for all possible sentences P, Q. The intent is to show that simply giving rules of inference for a piece of vocabulary is not sufficient to provide a meaningful piece of language. On this conclusion Nuel Belnap remarks:

> We must first, so the moral goes, have a notion of what *and* means, independently of the role it plays as premiss and as conclusion. Truth-tables are one way of specifying this antecedent meaning [...] (Belnap, 1962, pp. 130)

Requiring that truth-tables can be offered for connectives does prohibit terminology such as tonk[32] but such a requirement is essentially giving up on an inferential formulation of concepts. It would entail that, in Dummett's terminology, the central notion of the semantic theory is truth and thus be largely truth-conditional in spirit.

In order to defend the possibility of inferentially articulated connectives Belnap offers an alternative condition. Consider a language $\mathcal{L}$ and an extension $\mathcal{L}'$ constructed by adding a single connective $\circ$, together with inferential rules governing it, to $\mathcal{L}$. Belnap's constraint on possible connectives is that the extension is conservative in the following sense.[33]

[32]Consider that the introduction rules tell us that $P \vdash P$ tonk Q. Hence the row with P true and Q false must assign truth to P tonk Q. But the elimination rules tell us that any row assigning truth to P tonk Q must assign truth to both P and Q. Consequently there is no way to assign truth-values coherently to tonk.

[33]In addition to the requirement of conservativeness Belnap adds a restriction that a connective must be unique in the sense that if $\circ_1, \circ_2$ are connectives for which the same inferential rules are valid then we must have that $P \circ_1 Q \vdash P \circ_2 Q$ and $P \circ_2 Q \vdash P \circ_1 Q$. This condition prohibits for example connectives with only introduction rules and no elimination rules. For more detail see Belnap (1962).

Definition 7. Let $\vdash_{\mathcal{L}}$ and $\vdash_{\mathcal{L}'}$ be the consequence relations for $\mathcal{L}$ and $\mathcal{L}'$ respectively and let $P_1, ..., P_n, Q$ be $\mathcal{L}$-sentences. $\mathcal{L}'$ is a *conservative extension* of $\mathcal{L}$ if

$$P_1, ..., P_n \vdash_{\mathcal{L}'} Q \text{ only if } P_1, ..., P_n \vdash_{\mathcal{L}} Q$$

In a less formal formulation the extension is said to be conservative if the addition of new vocabulary does not add any new inferential relations between sentences containing only the old vocabulary. It can easily be shown that all standard logical connectives satisfy this property while, as demonstrated above, tonk does not.[34] Adding logical resources to a language is then a conservative extension which only allows for the expression of inferential relations already present.

Generalising this idea to not only deal with logical connectives and proofs Dummett offers the following definition[35]:

Definition 8. A language $\mathcal{L}$ is said to have *logical harmony* if for all sublanguages $\mathcal{L}'$ any material inference relation in the full language $\mathcal{L}$ between sentences in the sublanguage $\mathcal{L}'$ is already present in $\mathcal{L}'$ (Dummett, 1991, pp. 215-219).[36]

Dummett also gives a motivating example for his requirement of harmony for natural languages:

> A simple case would be that of a pejorative term, e.g. 'Boche'. The condition for applying the term to someone is that he is of German nationality; the consequences of its application are that he is barbarous and more prone to cruelty than other Europeans. We should envisage the connections in both directions as sufficiently tight as to be involved in the very meaning of the word: neither could be severed without altering its meaning. Someone who rejects the word does so because he does not want to permit a transition from the grounds for applying the term to the consequences of doing so. The addition of the term 'Boche' to a language which did not previously contain it would produce a non-conservative extension, i.e. one in which certain other statements which did not contain the term were inferable from other statements not containing it which were not previously inferable (Dummett, 1981, pp. 454).

[34] A proof for the standard connectives would entail showing that given a language with no logical symbols the extension by (for example) $\wedge, \neg$ is conservative. Having established this the conservativeness of all boolean connectives follows from Post's completeness theorem. A proof of a sufficient fragment of Post's theorem can be found as Corollary 1.6 of Mendelson (1997).

[35] This is not strictly correct. The condition offered is a *necessary condition* for logical harmony but not a sufficient one. Dummett doesn't offer a complete definition but rather a discussion which continues in the literature. For further reference see Dummett (1991) or Prawitz (2006).

[36] This is essentially the same definition as given by Belnap except formulated in terms of restricting to a sublanguage rather than extending it. For the reader familiar with category theory the definition simply says that all subcategories $\mathcal{L}'$ of $\mathcal{L}$ are full.

Hopefully we can all agree that 'Boche' is a defective concept. The point of requiring logical harmony of a language is to exclude concepts as this from being formed.

At this point we might ask if this requirement goes too far. It is not unreasonable to think that the rejection of prejudice, while admirable, should perhaps proceed on grounds not entirely devoid of empirical content. Adding fuel to the fire Brandom (1998, pp. 126-127) points out that the condition of logical harmony forces us to exclude even legitimate concepts. If we would require that our language had harmony then then we could never encounter new conceptual relations as they would entail a non-conservative extension of the language we already used. But novel conceptual content is central to all intellectual pursuits. In requiring logical harmony we would have to reject Einsteinian relativity in physics for not being a conservative extension of Newtonian physics. That a concept is not in harmony with our language simply means that it has some material, non-logical, content.[37] As a tentative solution Brandom instead offers that in encountering concepts which are not in harmony with the rest of our language we ought to discover the new material inferences it licenses and evaluate whether they are ones that we are entitled to. What makes certain concepts defective is that they would license inferences which, when made explicit by a conditional statement, we are not entitled to (Brandom, 1998, pp. 127-128).

4.2 Analyticity, Holism, and Compositionality

Perhaps the most renowned, and certainly the most spirited, criticism of inferentialism comes from Jerry Fodor and Ernie Lepore. In a pair of polemical papers (2001, 2007) they raise a family of familiar issues on meaning phrased in regards to inferential-role semantics. The first point they raise is the question of which inferences are to be considered constitutive of meaning. Unless the meaning of a sentence is to consist of all its inferential relations the theory of meaning needs to offer a criterion for which inferences are to be considered part of knowing what a sentence means. But such a distinction would divide the space of claims between those whose correctness we evaluate purely on the grounds of their meaning and those that we don't. It would, in essence, be a way of determining whether an inference, and hence the assertion of the corresponding conditional, is analytic or synthetic.

That this distinction can be maintained is one of the dogmata forcefully attacked by W.V. Quine. In his *Two Dogmas of Empiricism* (1951) he argues that any potential definition of the analytical is inescapably circular. Defining the analytical must rely on some previously available notion of synonymy which in turn would require us to already know what statements about synonymy are

[37] As an aside Brandom considers this to also be a criticism of truth-conditional semantics. Since they essentially offer as the meaning of a concept the kind of state of affairs that must hold for it to be applicable then it is not clear how we can talk about conceptual evolution rather than just the replacement of an old concept with a new one. In this way the requirement of logical harmony gives rise to an incommensurability problem in Kuhn's sense (1962, pp. 97-103).

analytical. Assuming that the inferentialist thinks there is some sufficient subcollection of inferences determining the meaning of a sentence they must then show that Quines problem be overcome. A possible inferentialist answer is given by Sellars in that what makes certain inferences analytic is that they are lawlike rather than accidental (1948, pp. 309-311). On this view what makes an inference analytic, and as such meaning-constitutive, is that it supports counterfactual reasoning about the concept. When attempting to distinguish analytical truths Quine agrees that they must be counterfactually stable by saying that a claim must be necessarily true in order to be analytical (Quine, 1951, pp. 28-30).[38] The problem then follows from Quine's view that we can't know what is necessary prior to knowing what is analytic. Fodor and Lepore (2007, pp. 183) take up this criticism that Sellars has reduced the original problem to explaining necessity rather than providing a solution. To this the inferentialist has a possible answer based on the notion of material inferences. Since we already know what inferences are considered materially valid according to our linguistic practice we also know what counterfactual relations our concepts stand in. Consequently there is no need to offer some criterion picking out the necessary inferences from the merely accidental; they are already given to us from out linguistic practice.

Grasping instead the other option available it's possible to claim that the meaning of a sentence is in fact constituted by all of the inferential relations it stands in. Fodor and Lepore finds this *prima facie* implausible since they don't see how any of us could then grasp concepts such as being made of iron without being experts on metallurgy (2001, pp. 470-471). Taking this path also entails a holism about meaning. If understanding a concept requires knowing all material inferences it figures in then to grasp a concept we need to also grasp all other concepts it is related to. According to Fodor and Lepore such a holism would be disasterous for our ability to communicate:

> How can we use the form of the words "It's raining" to communicate to you our belief that it's raining unless the word "raining" means the same to all of us? And, how can it mean the same to all of us if, on the other hand, no two people could conceivably agree on all the inferences in which "raining" occurs (to say nothing of the "correct inferences" in which it occurs)? (Fodor & Lepore, 2007, pp. 186)

That inferentialist semantics are at least somewhat holistic is not disputed by Brandom (2007, pp. 663). Instead he contends that holism is only a problem for understanding each other on a particular, in his terminology, Lockean view of communication (2010, pp. 333). On this view communication consists of the transportation of some mental content from one speaker to another. If communication is instead construed as succesfully making moves in a shared language-game with a scorekeeping practice then communication is possible between interlocutors even when they aren't both aware of the same inferential

[38]In Quine's discussion it even seems that 'Necessarily ...' becomes an operator that internalises in the language the claim that some assertion is analytic. That it has this function makes it a piece of logical vocabulary on the expressive view discussed above.

relations of the sentence uttered. The assertion that 'It's raining' is a move in the language game which changes the deontic score of everyone involved even if they don't entirely agree on how the score should be updated. Whether this gives an acceptable notion of communication is not immediately clear. Fodor and Lepore (2007, pp. 186) raise the issue that if we also disagree on the inferential relations of other relevant sentences then we wouldn't need to convey much similar content at all. Brandom concedes to this point but argues that by cooperating and making their inferential commitments explicit the interlocutors can inform each other of how they understand the assertion in question (Brandom, 1998, pp. 485-486).

Moving on from communication to the question of language production Fodor and Lepore raise the issue of compositionality. In order to explain how we can generate, and understand, an arbitrarily large amount of sentences with a finite mind it is generally agreed that language needs to be, at least somewhat, compositional (Lycan, 2008, pp. 110-111). Since meaning construed inferentially is, at least partially, holistic it seems to follow that the meaning of a sentence is not entirely determined by its parts. The fact that the meaning of subsentential expressions are to be understood through the meaning of the sentences they occur in makes inferential-role semantics, in Brandom's words, a "top-down" affair (2007, pp. 671). While conceding that the entire meaning of a sentence can't be determined by knowing just it's building blocks Brandom rejects the idea that this would prevent us from understanding and producing novel sentences. Encounters with an unknown sentence do however not occur in a linguistic vacuum. The compositionality of logical vocabulary, even when identified through its expressive function, allows us to reduce the meaning of complex sentences to that of simpler ones. Then our understanding of the substitution inferences that govern the subsentential expressions in the simple sentences allows us to recognise at least some of the material inferences the novel sentence occurs in. That our understanding of subsentential expression relies on our understanding of complete sentences is only a problem if we would require that we could understand a novel sentence without having ever encountered its components in other complete expressions. But such a condition is surely too strong. If, on the other hand, we're competent linguistic practitioners then when we encounter the new expression we already have access to the meanings of our subsentential expressions. We can then use that understanding to evaluate what the unfamiliar expression means.

Assuming that the inferentialist theory of meaning is correct it must be the case that language is not entirely compositional. Knowing the meaning (as inferential role) of each word that occurs in a sentence isn't in general sufficient to determine the meaning (as inferential role) of that sentence. From this it follows that the principle of compositionality, as Fodor and Lepore view it, is violated. As a result of this, and that they consider compositionality to be non-negotionable, they think that any theory that equates meaning with inferential role must be incorrect. On the other hand this objection isn't necessarily a defect of the theory. In fact Brandom thinks that it is a virtue:

> I agree with Fodor and Lepore that "in general the inferential role of a sentence/thought is not determined by the inferential roles of its constituents." But I think this fact evidences not a particular defect of inferentialism, but simply a fact about languages and (so) concepts. It is important not to treat languages as more compositional then they are. They are compositional with respect to their substitution inferences, but not with respect to the rest. (Brandom, 2007, pp. 675)

That meaning can be somewhat holistic and not entirely compositional is a point agreed on by Dummett. As an example of a social practice that is understandable and non-compositional he puts forward board games (1991, pp 222). In order to understand the significance of moves and pieces it is not enough to know just the rules governing that part. Knowing how the queen moves in a game of chess but not knowing what decides the victor would not determine the entire significance of the queen. On the other hand knowing the full rules of chess except for under what conditions *en passant* is allowed would not entail a complete ignorance of the game. For reasons such as this Brandom and Dummett think that linguistic practices occupy some middle position between the compositional and non-compositional. That inferential-role semantics turns out to have this property is, on this view, a feature rather than a a failing. Fodor and Lepore grant that this approach is viable but that it then seems to imply that there is some distinguished collection of inferential relations which determine the meaning of a sentence (2007, pp 190) which would again raise the question of the analytic/synthetic distinction. However this only follows if it is taken for granted that understanding novel sentences requires knowing its full meaning. Otherwise it is still possible to claim that while the meaning of a sentence does consist of all inferential relations it stands in we can still communicate even though our understanding of previously unfamiliar expressions is only partial. As an example consider a physicist talking to her father about when she will arrive home for the holidays. They manage to communicate intelligbly about time even though the physicist, being an expert in general relativity, has knowledge of inferences about the concept of time which her father has not.

4.3 Truth and Paradox

One possible advantage of an inferentialist approach to semantics is that it can deny the need for a theory of truth. Since, as we saw earlier, Brandom doesn't take this route his theory must must face the trial of semantic paradoxes related to truth. In particular he needs to have a story on how to deal with the Liar paradox. Being uncharacteristically brief Brandom defers, in an off-hand sentence, the matter entirely to a paper written by Dorothy Grover (1977). In it she argues that a sentence constructed through anaphoric means is only meaningful if the chain composed of its anaphoric antecedents is "grounded" (Grover, 1977, pp. 597-598). That an anaphoric expression is grounded means that there is some meaningful non-anaphoric expression that initates the chain of inheritance.[39] Given this criterion it follows that anaphoric expressions whose dependencies form a circle or infinite sequence must be considered devoid of content.

The approach that Grover takes to the Liar paradox then essentially consists of allowing for sentences to not have a truth-value. In particular this happens when the sentence in question contains an ungrounded anaphoric expression. Then she takes the Liar paradox to be a valid argument establishing that 'This sentence is false' is one such sentence without a truth-value as it would lead to contradiction if it had.

Considering defences such as this is what leads Kevin Scharp to introduce what he calls "revenge paradoxes" (2014, pp. 600). These are paradoxes which simply modify the original Liar-sentence to be a disjunction of truth and the new semantic vocabulary introduced to remove the paradox. In the case of Grover's approach such a modified sentence would be 'This sentence is false or has no truth-value'.[40] Working out the possibilities in the straightforward way this sentence leads to a contradiction. Hence even the prosentential theory falls prey to the problems of defining truth.

In response to this criticism Brandom accepts that truth can't be expressed in a theory strong enough to express the prosentential theory. He does however think this to be unproblematic since he never appeals to truth in the exposition of his theory of meaning (2010, pp. 357-359). He is not fazed by the fact that his theory of truth is excluded since it plays no important part in the project of explaining meaning.

This defense might seem promising but it quickly gets into trouble. Given the mechanisms of his semantic theory it is possible to construct a similar problem internal to it. In his discussion of truth as a "prosentence-forming operator" Brandom states:

> To take such a line is not to fall back into a subject-predicate picture, for there is all the difference in the world between a prosentence-forming operator and the predicates that form ordinary sentences. (Brandom, 1998, pp. 305)

This distinction is one that Brandom needs to maintain in order to avoid the internalisation of the paradox above. Keeping truth-assignment as an operator, a function mapping sentences to other sentences, means that for Brandom's theory it can get it's meaning through the expressive role it plays rather than through the inferential relations it occurs in. Consequently Scharps argument

[39] This criterion on anaphora is one that Brandom endorses (1998, pp. 458) and explicitly recognises as shared between Grover and Kripke (1998, pp. 322). The criterion of groundedness for anaphora may seem intuitively plausible but there are alternatives available. Peter Aczel (1988) has shown that by removing the analoguous requirement in set theory, that sets are well-founded, gives rise to a perfectly workable theory of sets and so of model theory. For a discussion on the Liar paradox from the vantage point of non-well-founded model theory see Barwise and Etchemendy (1987).

[40] This approach can also be found independently in Yanofsky (2003, pp. 10) coupled together with Quine's paradox making it instead rely on an indirect form of self-reference.

would then show that truth only can be considered as an operator and not a predicate since the latter would lead to paradox. Unfortunately it turns out that the distinction drawn is a difference which makes no difference.

To make the argument clear I will put it in formal terms. Let $\mathcal{S}$ be the collection of sentences and $\mathcal{T}$ the collection of terms in the language. Sentence nominalisation is then a function $\ulcorner\ \urcorner : S \to T$ constructing a term from a sentence.[41] Given the prosentence-forming operator '... is true', a mapping $\mathcal{S} \to \mathcal{S}$, then we can define a partial function T which maps the sentence nominalisation $\ulcorner P \urcorner$ to the sentence '$\ulcorner P \urcorner$ is true'. On Brandom's own account (1998, pp. 372) this is a predicate. Formally we have a partial function $T : \mathcal{T} \to \mathcal{S}$. Since the sentence outputted by '... is true.' inherits its content anaphorically we know that $T(\ulcorner P \urcorner)$ is a sentence with the same content as the sentence P. Having the same meaning, in Brandom's theory, is the same as standing in the same inferential relations. Then we can show[42] that for all sentences P that:

$$T(\ulcorner P \urcorner) \leftrightarrow P$$

This is precisely what it means for a language to have a "truth-predicate" in the sense forbidden by Tarski's theorem.[43]

Theorem. (A variant of Tarski's theorem.)[44] Let $\mathcal{L}$ be a sufficiently strong language. Then there is no predicate T in $\mathcal{L}$ such that

$$T(\ulcorner P \urcorner) \leftrightarrow P$$

By just using ingredients of Brandom's semantic theory this is a construction of a predicate whose existence leads to contradiction. The only extra assumption needed was that the prosentence-forming operator '... is true.' was present in the language.[45] Since the entire point of the discussion on truth was that Brandom wanted a theory of truth compatible with his semantics it seems clear that he

[41] As noted earlier Brandom's theory contains even more avenues for sentence nominalisation as he allows 'Goldbach's conjecture' to be the name of an entire collection of sentences. Since it is sufficient for the argument, and uniformly applicable to sentences, I will restrict myself to use the single quotation marks ' ' producing a term from a single sentence. For ease of reading I will write these marks as $\ulcorner\urcorner$.

[42] Write $\varphi \vdash \psi$ for the inferential relation from φ to ψ and assume that P, Q stand in the same inferential relations. Then $P \vdash R$ if and only if $Q \vdash R$. Clearly $P \vdash P$ and consequently $P \vdash Q$. In a similar fashion the converse can be shown. Since the role of $\to$ is to internalise valid inferences the result follows. This argument is a special case of the Yoneda lemma and can be applied in a more general categorical setting. For more details and a friendly introduction, see Riehl (2016).

[43] Tarski's theorem is generally proved in a context of formal languages with Gödel numberings. That path should be available to us here as well, through formalising Brandom's semantic commitments and justifying that natural numbers ought to be expressible in the object language, but would require too substantial a digression. Instead what follows is a variant of the theorem with a proof in Appendix A. For a more standard development of the theorem see Chapter 3 in Schwichtenberg and Wainer (2011).

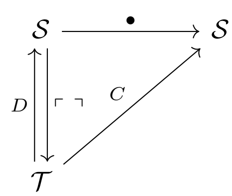
[44] For a more detailed formulation and proof see Appendix A.

[45] Actually not even this is strictly necessary. It is sufficent that the language contains truth-ascriptions from which we can extract the problematic predicate by substitution.

would accept that such locutions are present in the linguistic practices he wished to explain. An option still open to Brandom at this point is to save the theory by rejecting that there is any meaning at all to assertions of truth which would block the construction of the problematic predicate.

The core of the argument above comes from the failure to maintain a distinction between the prosentence-forming operator and a predicate. In a language which allows for nominalisation and disquotation of sentences such a distinction is simply not available. Since there is a systematic, and invertible, way to produce terms from sentences we can define a predicate which takes a term (of the form $t = \ulcorner S \urcorner$ for some sentence S) applies disquotation and then applies the operator to the resulting sentence. This predicate would only be defined on those terms which are sentence nominalisations but there is no reason to require all predicates to be total. In the other direction given a predicate we can define an operator by composing the predicate with nominalisation.

In order to clarify the argument consider the diagram:



where $\bullet$ is an operator, C a predicate, and D being disquotation[46]. Note that an operator is simply a function from the collection $\mathcal{S}$ of sentences to itself. If we are given an arbitrary operator $\bullet$ then we can define a predicate C by

$$C(t) \stackrel{\text{def}}{=} \bullet D(t)$$

Being defined only on those terms which are sentence nominalisations we know that $t = \ulcorner P \urcorner$ for some sentence P. Hence we can rewrite the predicate as

$$C(\ulcorner P \urcorner) = \bullet D(\ulcorner P \urcorner) = \bullet P$$

In the other direction consider a given predicate C defined on the terms $t = \ulcorner P \urcorner$ for some sentence P. Then there is no trouble constructing an operator $\bullet$ by defining it as

$$\bullet P \stackrel{\text{def}}{=} C(\ulcorner P \urcorner)$$

Since C is defined for all terms $t = \ulcorner P \urcorner$ it follows that $\bullet$ is defined for all sentences P.

These arguments tell us that operators and predicates defined on nominalised sentences are in a bijective correspondence. By applying nominalisation or disquotation we can construct an object of the other type which figures in exactly the same inferential relations as the original.[47]

[46] The disquotation function D simply defined as the left-inverse of $\ulcorner\ \urcorner$. This means that $D(\ulcorner P \urcorner) = P$ for all sentences P.

[47] Considered as sets, or objects in a category with the relevant constructions, this result is

5 Expressive Completeness

In the preceeding section Brandom's theory of truth ran into trouble because it could not be expressed inside the language to which it's supposed to be applicable. Additionally I made note of the fact that using only the apparatus of his semantic theory and that everyday truth-ascriptions are meaningful it was possible to construct a problematic theory of truth. What got his theory in trouble is Brandom's committment to *expressive completeness*. As he puts it:

> How much logical vocabulary is worth reconstructing in this fashion? In this project, neither more nor less than is required to make explicit within the language the deontic scorekeeping social practices that suffice to confer conceptual contents on nonlogical vocabulary in general. (Brandom, 1998, pp. xx)

Remember that Brandom's conception of logical vocabulary is that it allows for expressing in the language what is implicit in our linguistic practice. This means that "logical vocabulary" here includes the vocabulary required for formulating his semantic theory. What Brandom wants is a theory of meaning which requires no metatheory to state or apply because such a theory wouldn't itself stand in need of explanation. The criterion of expressive completeness might then seem highly desireable if difficult to spell out. Nevertheless it will turn out to be unacheivable. To show this I need to introduce a necessary condition for expressive completeness which we will see can't be satisfied by Brandom's theory.

The key notion, due to Scharp[48], for the argument is *internalisability*. Scharp maintains the distinction, attributed above to Dummett, between a theory of meaning and a semantic theory. That a semantic theory can be internalised formalises the intuitive idea that it can be expressed within the language it applies to.

Definition 9. A semantic theory T is *internalisable* in a language $\mathcal{L}$ if it has an extension $\mathcal{L}'$ such that every claim of T can be translated to a sentence in $\mathcal{L}'$ and T correctly specifies the meanings of all sentences in $\mathcal{L}'$. (Scharp, 2010, pp. 267)

Since my goal is to produce a *reductio*[49] there is no need to digress on internalising theories of meaning. It's sufficient to say that if a theory of meaning is internalisable in a language[50] then so is some semantic theory of the kind it describes. Otherwise there would be no assignment of contents to the constituent

no surprise. Given that $\ulcorner\ \urcorner$ has a left-inverse its image $\ulcorner\mathcal{S}\urcorner$ (which consists of the predicates defined on nominalised sentences) is isomorphic to $\mathcal{S}$. Consequently this gives rise to another isomorphism between the objects $\mathcal{S}^{\mathcal{S}}$ and $\mathcal{S}^{\ulcorner\mathcal{S}\urcorner}$ by the Yoneda lemma.

[48] The exposition here follows the simplified version of Scharps theory of internalisability as put forward in (Scharp, 2010). For a more detailed discussion see Scharp (2014).

[49] In the intuitionistically aboveboard sense of negation introduction and not requiring the Law of the Excluded Middle.

[50] This terminology differs slightly from Scharps in that it makes internalisability of a theory of meaning into a relation with a language. Scharp instead takes the language that theory is stated in as basic and expresses the criteria in terms of it.

sentences of the semantic theory which coheres with the precepts of the theory of meaning.

That a theory of meaning is expressively complete requires that it can be internalised in a language it applies to in order for it to be used to explain linguistic meaning without leaving unanswered questions of it's own, constituent sentences, meaning. This in turn requires that a compatible semantic theory can be internalised in that same language.

In the light of these notions Scharps argument, expounded on in the previous section, was that Brandom's theory of truth can't be expressed in the language it applies to and so, if his semantic theory should assign meaning to it, neither can his theory of meaning. To defend his theory of meaning Brandom claimed that his theory of truth was an outside component which needn't be internalised. In response to that defense I offered a modified version of Scharp's argument as follows:

(1) If a theory of meaning is expressively complete then it is internalisable in a language it applies to.

(2) If the theory of meaning is internalisable in a language then so is a semantic theory, compatible with that theory of meaning, for that language.

(3) If a semantic theory, compatible with Brandom's theory of meaning, is internalisable in a language it applies to and that linguistic practice contains meaningful truth-ascriptions then that language contains a predicate T such that

$$T(\ulcorner P \urcorner) \leftrightarrow P$$

for all sentences P.

(4) A variant of Tarski's theorem, contradicting the existence of such a predicate T, is provable in a language containing Brandom's semantic theory.

Therefore, given the assumptions that our linguistic practice contains ascriptions of truth and that Brandom's theory of meaning is expressively complete, we can prove a contradiction.[51] To accomodate this finding there are then two main options. Brandom has the option to reject outright the need for a theory of truth and simply take our everyday talk of truth to be defective. He could also follow Tarski and accept the need for a metalanguage in which to state the semantic theory (Tarski, 1944, pp. 348-351) but this path would entail giving up on expressive completeness.

At this point it might seem that the prudent option is to reject truth and remain content that the theory is expressively complete. It turns out that this is not a real option since a modified version of the argument replacing the role of truth still goes through. Before proceeding I do need to state a lemma already implicitly used in proving Tarski's theorem.

[51]The statements (3) and (4) are what was shown in the previous section.

Lemma. (Diagonal Lemma)[52] Let P be a predicate defined on all nominalised sentences. Then there is a sentence Q such that

$$P(\ulcorner Q \urcorner) \leftrightarrow Q$$

With the lemma in hand the next step is to consider the pragmatics of assertion. For example:

$$\text{Giraffes are the silliest looking animals.} \tag{3}$$

Being an assertion means that part of its force is committing the speaker to it's content. In order to make this explicit our linguistic practices contains the (in Brandom's expressive sense) logical connective '... is committed to ...' and in particular 'I am committed to ...'. This allows us to make assertible the force that was implicit in the above statement by saying:

$$\text{I am committed to 'Giraffes are the silliest looking animals.'.} \tag{4}$$

Thanks to the role of expressive vocabulary the force of this expression is the same as that of an asserting (3). As we saw earlier no clear distinction can be drawn between predicates and logical operators in such a rich theory and so we can make this expression into a predicate C. This C is a predicate which maps a sentence nominalisation $\ulcorner P \urcorner$ to the sentence 'I am committed to $\ulcorner P \urcorner$'. Appending negation provides a predicate $\neg C$ to which we can apply the Diagonal lemma to produce a sentence Q such that:

$$\neg C(\ulcorner Q \urcorner) \leftrightarrow Q$$

Since these statements are equivalent, in the expressive sense codifying that each can be inferred from the other, they must have the same meaning.[53] On Brandom's conception the force of an expression is determined entirely by the sense of the sentences and the type of speech-act involved. As they have the same sense it then follows that the force of asserting Q and asserting $\neg C(\ulcorner Q \urcorner)$ is the same. But as we saw above the force of assering Q is the same as the force of asserting $C(\ulcorner Q \urcorner)$. This leads us to the conclusion that asserting and denying committment to Q has the same force. Then whenever someone asserts Q their interlocutors both ought and ought not to attribute to the speaker commitment Q. Such a statement can't be coherently assigned any sense in an inferentially formulated semantics and yet it is perfectly constructable on the assumption that commitment is internally expressible in the language.

This conclusion cuts off the option of saving the expressive completeness of Brandom's theory by denying the need for a theory of truth. Commitment plays a central role for Brandom and if it can't be internalised then neither can his theory of meaning.

[52] Once again see Appendix A for a more detailed discussion and proof.

[53] Writing $\vdash$ for the inferential relation we know that if $A \leftrightarrow B$ then $A \vdash B$ and $B \vdash A$. If $A \vdash C$ or $C \vdash A$ it follows that $B \vdash C$ or $C \vdash B$ respectively. Symmetrically the result is the same if $B \vdash C$ or $C \vdash B$. Then A and B are entirely intersubstitutable in inferences and so have the same meaning.

6 A General Restriction on Theories of Meaning

In the previous sections I have shown how Brandom's theory of meaning runs into several problems. The central theme of these objections has been that a large amount of self-reference sits ill at ease with the objective of expressive completeness. Each of the arguments presented have made use of some attribute of truth or idiosyncracy of Brandom's theory and consequently could be thought to be particular to it. The aim of this section is to show that this is not the case by singling out the core features of the arguments above.

Before proceeding too far into this section I want to collect all the pieces of notation that will be used. I will, as previously in this text, work in a primitive functional view on the building blocks of language. We start with a collection of terms $\mathcal{T}$ and a collection of complete sentences $\mathcal{S}$. From these we can generate a collection of predicates $\mathcal{S}^{\mathcal{T}}$ as functions taking a term and producing a sentence. The language we are working in also contains a negation operator $\neg$.

At the most essential level a semantic theory is a collection of statements about the building blocks of language whichever they may be. Whatever the central notion of a particular semantic theory that theory is in the business of making claims about sentences and predicates. If we have a particular semantic theory for a language $\mathcal{L}$ it's reasonable to ask about the most basic prerequisites that $\mathcal{L}$ must satisfy in order to internalise that semantic theory. As some of the terms of the semantic theory are predicates and sentences of $\mathcal{L}$ the language must have a robust ability for nominalisation.

This should come as no surprise. Being able to talk about sentences and predicates as if they are objects themselves is something we're entirely used to. Taking a sentence and turning it into a noun requires only that we apply quotation marks around it. The use of quotation marks is also what gives name to the inverse process of *disquotation* taking a nominalised noun and producing again the original sentence. In order to internalise a semantic theory these are indispensible tools. Treating these processes as functions on our grammatical collections we can write them as:

$$\mathcal{T} \underset{\ulcorner\;\urcorner}{\overset{D}{\rightleftarrows}} \mathcal{S}$$

Here D stands for disquotation. For reasons of legibility I use $\ulcorner\;\urcorner$ insted of quotation marks. Additionally we can note that $D(\ulcorner S \urcorner) = S$ for all sentences S since first applying and then removing quotation marks makes no difference. Nominalisation and disquotation of predicates works in an entirely parallel fashion:

$$\mathcal{T} \underset{\ulcorner\;\urcorner}{\overset{D}{\rightleftarrows}} \mathcal{S}^{\mathcal{T}}$$

In a slight abuse of notation I will use the same symbols for this case. It is entirely conceivable to have a language with only a restricted amount of nominalisation. That language would then only allow for some of it's sentences and predicates to be treated as nouns. However such a language could not internalise it's own semantic theory as there would be sentences which the semantic theory makes claims about but that the language itself can't apply predicates to.

The next step is to borrow a very powerful mathematical theorem: Lawvere's Fixed Point theorem. It can be considered an abstract version of the Diagonal lemma appealed to in a previous section.[54] In the statement of it there are two pieces of technical vocabulary: Representing all functions and being a fixed point of a function. An intuitive understanding of these should be sufficient to follow the argument but for the reader craving specifics the details can be found in Appendix B.

Theorem. (Lawvere's Fixed Point Theorem)[55] Let $f : A \to B^A$ be a function which represents all functions $h : A \to B$. Then all functions $\alpha : B \to B$ have a fixed point.

With this machinery in place it's possible to state and prove the main result of this section:

Theorem. No language containing negation can allow for unrestricted nominalisation.

Proof. Let $\mathcal{L}$ be a language with a negation operator $\neg : \mathcal{S} \to \mathcal{S}$ and a pair of nominalisation and disquotation operators

$$\mathcal{T} \underset{\ulcorner\ \urcorner}{\overset{D}{\rightleftarrows}} \mathcal{S}^{\mathcal{T}}$$

Let $P \in \mathcal{S}^{\mathcal{T}}$ be an arbitrary predicate. Then $D(\ulcorner P \urcorner) = P$ which means that D represents all functions from $\mathcal{T} \to \mathcal{S}$. Hence, by Lawvere's theorem, all functions $\alpha : \mathcal{S} \to \mathcal{S}$ in $\mathcal{L}$ have a fixed point. But this contradicts the fact that

$$\neg S \not\leftrightarrow S$$

for all sentences S. □

As noted in the buildup to this argument any language capable of internalising it's semantic theory would have to be capable of unrestricted nominalisation. Consequently no language which contains negation[56] can internalise it's own semantics or, indeed, it's theory of meaning. Since any theory of meaning that wishes to explain everyday linguistic practices would need to be applicable to such languages it is clear that it can't be expressively complete.

[54]In fact it generalises not only the Diagonal lemma but also, among others, the Gödel's Incompleteness theorems, the Halting problem, Cantor's theorem, and Russell's paradox. For exposition and discussion on this see Yanofsky (2003).

[55]The theorem is stated here in the language of sets but that is purely a convenience. For more detail and a proof see Appendix B.

[56]There are two reasons that negation plays a part in this argument; it has no fixed points and it is a central part of everyday linguistic practice. Any operator on sentences which has these properties could fill the same role in this argument and could also have been used instead of negation in the previous sections.

7 Summary and Conclusion

The road traveled through this text has been one of ever stronger limits to what a single language can express. First we saw Scharp's argument that in trying to avoid the Tarskian limitations to internalising truth we run afoul of new contradictions. To make matters worse, the method he uses doesn't rely on any details of attempted solutions and so generalises easily to any such attempt. In response to this Brandom attempts to salvage the prosentential theory by keeping truth as a logical operator (in the expressive sense) that Scharp's construction doesn't apply to. As a counterpoint I showed that this distinction is a difference that doesn't make a difference in any language rich enough for Brandom's theory. Moving on from talk of truth to talk of committment I showed that problems of internalisation plague this semantic device as well. In the case of committment what is produced isn't a logical contradiction but instead a practical one. The statement constructed is one we can't possibly perform the pragmatics of. Since Brandom's theory of meaning has committment as a central concept it can't then be internalised in a language it applies to. What was required as premises for both these arguments was however only that the language had a sufficient ability to internalise talk of its own sentences and some operator which couldn't possibly preserve the meaning of sentences it was applied to. The first of these criteria is necessary in order for a language to be able to internalise its theory of meaning. The second criteria is in turn fulfilled by all languages containing logical negation. Consequently these arguments generalise to prohibit in general the internalisation of a theory of meaning for a language sufficiently strong to be interesting.

Throughout these arguments I have refrained from stating the paradoxical sentences in the form:

$$\text{(5) is false.} \tag{5}$$

The reason for this is that the straightforward way of proving that a contradiction follows uses the classical Law of the Excluded Middle or bivalence to check all the possible cases. I have chosen to state them more formally here to make explicit the well-known fact that the results do not depend essentially on any such principle.

An important question for philosophy of language is this: Can any single theory of meaning be applicable to all of human language? The search for such a theory is a noble ambition since finding one would allow us to finally put to rest age-old questions of meaning. Unfortunately the results of this text would point towards a negatory answer. If it is indeed the case that no single theory of meaning can do the job, the possibility remains open that multiple such theories can. The phenomena that needs saving is our ability to communicate and express ourselves through language and not the existence of some singular well-defined meaning for every particular utterance.

A Tarski's Theorem and the Diagonal Lemma

Before I can state a proof of the theorem I need to clarify what it means for a language to be "sufficiently strong". The first criterion is that it contains negation. Secondly the language must allow for sentence and predicate nominalisation in the sense that there exists functions $f : \mathcal{S} \to \mathcal{T}$ and $g : \mathcal{S}^{\mathcal{T}} \to \mathcal{T}$ where $\mathcal{S}$ is the collection of complete sentences, $\mathcal{T}$ is the collection of terms, and $\mathcal{S}^{\mathcal{T}}$ is the collection of, unary, predicates. The function f takes in a sentence and produces a term that names it. The function g instead takes in a predicate and outputs a term that names that. The only difference between the functions is their domain of definition. For convenience I will use the notation $\ulcorner\ \urcorner$ for both these functions.[57]

That Brandom's theory does satisfy these critera is readily clear from his discussion of negation as the minimal incompatible statement as well as the role of nominalisation within his own theory of truth. As a final prerequisite I will prove the following lemma:

Lemma. (Diagonal Lemma) Let $\mathcal{L}$ be a language with nominalisation and let P be a predicate defined on all nominalised sentences. Then there is a sentence Q such that

$$P(\ulcorner Q \urcorner) \leftrightarrow Q$$

Proof. Let $R : \mathcal{T} \to \mathcal{S}$ be an arbitrary predicate, defined on all nominalised sentences, and let $f : \mathcal{T} \to \mathcal{S}$ be the predicate defined by

$$f(\ulcorner R \urcorner) \overset{\text{def}}{=} \forall y(y \doteq \ulcorner R(\ulcorner R \urcorner)\urcorner \to P(y))$$

I will show that

$$f(\ulcorner R \urcorner) \leftrightarrow P(\ulcorner R(\ulcorner R \urcorner)\urcorner)$$

First I deal with the rightwards direction:

$$\dfrac{\dfrac{\dfrac{f(\ulcorner R \urcorner)}{\forall y(y \doteq \ulcorner R(\ulcorner R \urcorner)\urcorner \to P(y))}}{(\ulcorner R(\ulcorner R \urcorner)\urcorner \doteq \ulcorner R(\ulcorner R \urcorner)\urcorner \to P(\ulcorner R(\ulcorner R \urcorner)\urcorner)}\,\forall E \qquad \ulcorner R(\ulcorner R \urcorner)\urcorner \doteq \ulcorner R(\ulcorner R \urcorner)\urcorner}{P(\ulcorner R(\ulcorner R \urcorner)\urcorner)}\,\to E$$

For the other direction there is the proof:

$$\dfrac{\dfrac{\dfrac{\dfrac{[y \doteq \ulcorner R(\ulcorner R \urcorner)\urcorner]^1 \qquad P(\ulcorner R(\ulcorner R \urcorner)\urcorner)}{P(y)}\,S}{y \doteq \ulcorner R(\ulcorner R \urcorner)\urcorner \to P(y)}\,\to I_1}{\forall y(y \doteq \ulcorner R(\ulcorner R \urcorner)\urcorner \to P(y))}\,\forall I}{f(\ulcorner R \urcorner)}$$

[57]This mild abuse of notation can be justified by taking the disjoint union of $\mathcal{S}$ and $\mathcal{S}^{\mathcal{T}}$ as the domain of $\ulcorner\urcorner$. In that case the nominalisation function is uniquely determined by a piecewise definition.

Now substitute in f for R and which produces:

$$f(\ulcorner f \urcorner) \leftrightarrow P(\ulcorner f(\ulcorner f \urcorner) \urcorner)$$

Letting $Q = f(\ulcorner f \urcorner)$ and tidying up the result is:

$$Q \leftrightarrow P(\ulcorner Q \urcorner)$$

□

With these prerequisites in hand we can restate the theorem in detail and produce a proof.

Theorem. (Undefinability of Truth) Let $\mathcal{L}$ be a language with negation and nominalisation. Then there is no predicate T in $\mathcal{L}$, defined on all sentence nominalisations, such that

$$T(\ulcorner P \urcorner) \leftrightarrow P$$

for all sentences P.

Proof. Let T be a predicate as in the statement of the theorem. Then since the language has negation it's possible to define the predicate $F : \mathcal{T} \to \mathcal{S}$ by $F(t) \stackrel{\text{def}}{=} \neg T(t)$. This predicate, like T, is only defined for those terms that are nominalisations of sentences. Since $\mathcal{L}$ has nominalisation we know that the Diagonal lemma holds for it. Applying it produces a sentence Q such that

$$F(\ulcorner Q \urcorner) \leftrightarrow Q$$

But from the definitions of F and T it follows that

$$Q \leftrightarrow F(\ulcorner Q \urcorner) \leftrightarrow \neg T(\ulcorner Q \urcorner) \leftrightarrow \neg Q$$

which is a contradiction. Hence T can't exist. □

B Lawvere's Fixed Point Theorem

The theorem proven in this section is generally shown in a category theoretic context. However in order not to get bogged down in the categorial details the proof will be performed in a similar framework to that of Yanofsky (2003).[58] All the objects involved will be construed as sets and functions in the standard framework of Zermelo-Fraenkel set theory but readers suspicious of sets can be reassured that this is only to help intuition. The argument employed here works, and was originally given by Lawvere (1969), in a general categorial setting.

In order to set the stage for the theorem I will need to introduce a few notions and notations. $A \times B$ is called a *cartesian product* and stands for the set of pairs (a, b) where a is an element of A and b an element of B. If we have functions $f : A \to X$ and $g : B \to Y$ then we can define the *product function* $f \times g$ from $A \times B$ to $X \times Y$ by:

$$f \times g(a, b) = (f(a), g(b))$$

The collection denoted by B^A is called a *function space* and is the set of all functions from A to B. The function $ev : B^A \times A \to B$ is the *evaluation function* defined by:

$$ev(f, a) = f(a)$$

In addition to these pieces of notation there are two technical terms to define. First up is *representation*.

Definition 10. A function $f : X \to Z^Y$ *represents all functions* from Y to Z if for each function $g : Y \to Z$ there is an element x in X such that $f(x) = g$.

The idea here is that the elements of X works as names for the functions $g : Y \to Z$ and that f simply takes in a name x and produces the outputs the function that x is the name of. The definition then states that every function in Z^Y is named by some element of x through f.[59]

Definition 11. A *fixed point* of a function $g : X \to X$ is an element x of X such that $g(x) = x$.

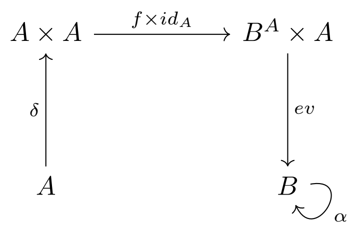
With all this in hand we can restate the theorem and provide a full proof.

Theorem. (Lawvere's Fixed Point Theorem) Let $f : A \to B^A$ be a function which represents all functions $h : A \to B$. Then all functions $\alpha : B \to B$ have a fixed point.

[58] In Yanofsky's paper the theorem is called the "Diagonal Theorem" because of it's similarity with the Diagonal Lemma. In order to avoid confusion I will instead use the, also standard, name "Lawvere's Fixed Point Theorem".

[59] This requirement, when stated in the language of sets, is simply that f is a surjection. However surjectivity is notoriously hard to generalise properly to a categorial setting as neither epimorphisms, split-epimorphisms, or point-surjections quite capture the same intuition. For this reason I've chosen to formulate it in terms of representation.

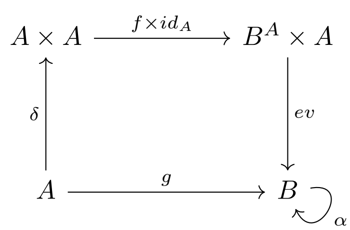
Proof. Consider the following diagram:



The functions δ and id_A are defined as $\delta(a) = (a, a)$ and $id_A(a) = a$. By defining the function $g : A \to B$ as

$$g(a) = \alpha(ev(f(a), id_A(a)))$$

we can fill in the diagram so that both paths around it are the same.



Since f represents all functions from A to B it in particular represents g. This means that there is $a_g \in A$ such that $f(a_g) = g$. Then

$$\begin{aligned} g(a_g) &= \alpha(ev(f(a_g), id_A(a_g))) \\ &= \alpha(ev(g, a_g)) \\ &= \alpha(g(a_g)) \end{aligned}$$

Consequently $g(a_g)$ is a fixed point for α. $\square$

References

[1] Aczel, P. (1988) *Non-well-founded sets*. Stanford: Center for the Study of Language and Information.

[2] Awodey, S. (2014) *Univalence as a Principle of Logic*, lecture notes, Calgary Mathematics and Philosophy Lecture Series, 2 October, Calgary.

[3] Barwise, J. and Etchemendy, J. (1987) *The Liar: An Essay on Truth and Circularity*. New York: Oxford University Press.

[4] Belnap, N. (1962) Tonk, Plonk and Plink. *Analysis*, vol. 22, no. 6, pp. 130-134.

[5] Brandom, R. (1998) *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Revised ed. Cambridge: Harvard University Press.

[6] Brandom, R. (2000) *Articulating Reasons: An Introduction to Inferentialism*. Cambridge: Harvard University Press.

[7] Brandom, R. (2007) Inferentialism and Some of Its Challenges. *Philosophy and Phenomenological Research*, vol. 74, no. 3, pp. 651-676.

[8] Brandom, R. (2010) Reply to Jerry Fodor and Ernest Lepore's "Brandom Beleaguered". In Weiss, B. and Wanderer, J. (eds.) *Reading Brandom: On Making It Explicit*. Routledge: New York, pp. 332-337.

[9] Brandom, R. (2010) Reply to Kevin Scharp's "Truth and Expressive Completeness". In Weiss, B. and Wanderer, J. (eds.) *Reading Brandom: On Making It Explicit*. Routledge: New York, pp. 357-359.

[10] Davidson, D. (1963) Actions, Reasons, and Causes. *The Journal of Philosophy*, vol. 60, no. 23, pp. 685-700.

[11] Davidson, D. (1967). Truth and Meaning. *Synthese*, vol. 17, no. 3, pp. 304-323.

[12] Dummett, M. (1981) *Frege: Philosophy of Language*. 2nd ed. Cambridge: Harvard University Press.

[13] Dummett, M. (1991). *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.

[14] Fodor, J. and Lepore, E. (2001) Brandom's Burdens: Compositionality and Inferentialism. *Philosophy and Phenomenological Research*, vol. 63, no. 2, pp. 465-481.

[15] Fodor, J. and Lepore, E. (2007) Brandom Beleaguered. *Philosophy and Phenomenological Research*, vol. 74, no. 3, pp. 677-691.

[16] Gentzen, G. (1934) Investigations into Logical Deduction. In Szabo, M. E. (ed.) (1969) *The Collected Papers of Gerhard Gentzen*. Amsterdam: North-Holland Pub. Co.

[17] Grover, D., Camp, J, and Belnap, N. (1975) A Prosentential Theory of Truth. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 27, no. 2, pp. 73-125.

[18] Grover, D. (1977) Inheritors and Paradox. *The Journal of Philosophy*, vol. 74, no. 10, pp. 590-604.

[19] Kuhn, T. (1996) *The Structure of Scientific Revolutions*. 3rd ed. Chicago: The University of Chicago Press.

[20] Lawvere, F. W. (1969) Diagonal Arguments and Cartesian Closed Categories. *Lecture Notes in Mathematics*, vol. 92, pp. 134-145.

[21] Lewis, D. (1970) General Semantics. *Synthese*, vol. 22, no. 1-2, pp. 18-67.

[22] Lewis, D. (1979) Scorekeeping in a Language Game. *Journal of Philosophical Logic*, vol. 8, no. 1, pp. 339-359.

[23] Lycan, W.G. (2008) *Philosophy of Language: A Contemporary Introduction*. 2nd ed. New York: Routledge.

[24] Mendelson, E. (1997) *Introduction to Mathematical Logic*. 4th ed. London: Chapman & Hall.

[25] Prawitz, D. (1974) On the Idea of a General Proof Theory. *Synthese*, vol. 27, no. 1-2, pp. 63-77.

[26] Prawitz, D. (2006) Meaning Approached via Proofs. *Synthese*, vol. 148, no. 3, pp. 507-524.

[27] Prior, A.N. (1960) The Runabout Inference-Ticket. *Analysis*, vol. 21, no. 2, pp. 38-39.

[28] Quine, W.V. (1951) Two Dogmas of Empiricism. *The Philosophical Review*, vol. 60, no. 1, pp. 20-43.

[29] Riehl, E. (2016) *Category Theory in Context*. New York: Dover Publications.

[30] Scharp, K. (2010) Truth and Expressive Completeness. In Weiss, B. and Wanderer, J. (eds.) *Reading Brandom: On Making It Explicit*. New York: Routledge, pp. 262-275.

[31] Scharp, K. (2014) Truth, Revenge, and Internalizability. *Erkenntnis*, vol. 70, sup. 3, pp. 597-645.

[32] Schwichtenberg, H. and Wainer, S.S. (2011) *Proofs and Computations*. Cambridge: Cambridge University Press.

[33] Sellars, W. (1948) Concepts as Involving Laws and Inconceivable without Them. *Philosophy of Science*, vol. 15, no. 4, pp. 287-315.

[34] Sellars, W. (1953) Inference and Meaning. *Mind*, vol. 62, no. 247, pp. 313-338.

[35] Sellars, W. (1957) Counterfactuals, Dispositions, and the Causal Modalities. *Minnesota Studies in the Philosophy of Science*, vol. 2, pp. 225-308.

[36] Soames, S. (1992). Truth, Meaning, and Understanding. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 65, no. 1-2, pp.17-35.

[37] Speaks, J. (2006) Truth Theories, Translation Manuals, and Theories of Meaning. *Linguistics & Philosophy*, vol. 29, no. 4, pp.487-505.

[38] Tarski, A. (1944) The Semantic Conception of Truth: And the Foundation of Semantics. *Philosophy and Phenomenological Research*, vol. 4, no. 3, pp. 341-376.

[39] Troelstra, A. S. and Schwichtenberg, H. (2000) *Basic Proof Theory*. 2nd ed. Cambridge: Cambridge University Press.

[40] Yanofsky, N. S. (2003) A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points. *Bulletin of Symbolic Logic*, vol. 9, no. 3, pp. 362-386.