Inquiry
An Interdisciplinary Journal of Philosophy

The Concept of Truth and the Semantics of the Truth Predicate

Emil Badici [a]; Kirk Ludwig [b]
[a] Louisiana State University, USA
[b] University of Florida, USA

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

The Concept of Truth and the Semantics of the Truth Predicate

EMIL BADICI* and KIRK LUDWIG**

*Louisiana State University, USA; **University of Florida, USA*

ABSTRACT   *We sketch an account according to which the semantic concepts themselves are not pathological and the pathologies that attend the semantic predicates arise because of the intention to impose on them a role they cannot fulfill, that of expressing semantic concepts for a language that includes them. We provide a simplified model of the account and argue in its light that (i) a consequence is that our meaning intentions are unsuccessful, and such semantic predicates fail to express any concept, and that (ii) in light of this it is incorrect to characterize the pathology simply as semantic inconsistency; a more nuanced view of the problem is needed. We also show that the defects of the semantic predicates need not undercut the use of a truth theory in a compositional semantics for a language containing them because the meaning theory per se need not involve commitment to the axioms of the truth theory it exploits.*

There are broadly speaking two different attitudes toward the family of semantic paradoxes of which the Liar, exemplified by (L), is a paradigmatic example.

(L)   (L) is false.

(L) appears to be a well-formed, fully meaningful, declarative sentence of English. It seems then that it must be either true or false. But each assumption leads to the opposite, and we are thus led to the conclusion,

(L) is false iff (L) is true,

Correspondence Address: Emil Badici, Philosophy Department, Louisiana State University, 106 Coates Hall, Baton Rouge, LA 70803-3901, USA. Email: badici@lsu.edu; Kirk Ludwig, Philosophy Department, University of Florida, Gainesville, FL 32611-8545, USA. Email: ludwig@ufl.edu

which, we feel like saying, cannot be true. According to the first attitude, the problem is only apparent, and the job of the philosopher is to find a mistaken assumption or step in the reasoning which absolves the semantic predicates of guilt. We will put this by saying that the idea is to find some way in which the reasoning misrepresents our meaning practices in using the expressions involved in the reasoning.[1] According to the second attitude, the reasoning reflects accurately our meaning practices with regard to the expressions which appear in the argumentation, and consequently there is a sense in which the relevant terms are genuinely pathological. Within this camp there are two further subdivisions. On the first, we accept the results[2] and admit that a contradiction can be true—and false—and we deal with the consequences. On the second, we retreat to providing an account of what has gone wrong and how best to respond. In the following, we develop some ideas connected with an account of this last sort.

We believe that there are genuine pathologies in the meaning practices associated with semantic predicates in natural languages. However, in our view, the semantic concepts themselves are not pathological. The pathologies that attend the semantic predicates arise because of the intention to impose on them a role they cannot fulfill, that of expressing semantic concepts for a language that includes them. A consequence is that our meaning intentions are unsuccessful, and such semantic predicates fail to express the concepts they are intended to.[3] This has consequences for how we should understand the nature of the pathologies in our meaning practices and how we should deal with them.

Our first goal in this paper is to provide a plausibility argument for the thesis that semantic predicates in English (and other natural languages) exhibit a kind of pathology that arises out of the intention to impose on them the role of expressing a certain semantic concept we grasp independently, which, though it might have seemed antecedently plausible that they could fulfill, they demonstrably cannot, and then to use this perspective to say something about the kind of pathology involved. Our second goal is to investigate the implications of this for truth-theoretic semantics for natural languages.

In section I, we develop in a simplified setting a model of the diagnostic account we have in mind. The guiding idea is that we can think coherently about truth and falsity in relation to sentences and utterances of sentences in natural language and that the difficulty arises in the natural attempt to express what we can think in the language we think about. The model and general considerations make this account plausible. The model also helps to shed light on the status of the pathologies involving semantic predicates and to show why ordinary communication about truth is not impeded despite the truth predicate not fulfilling its intended role. In section II, we consider what this means for the possibility of giving a truth-theoretic semantics for natural languages. We argue, in line with earlier work by Ludwig (2002) and

Lepore and Ludwig (2005), that proper attention to the content of such a theory shows that the pathologies of the semantic predicates do not prevent us from providing a semantic theory for the language because, though a truth-theoretic semantics for the whole of the language, including its pathological predicates, cannot be given in a metalanguage which does not itself contain suspect predicates, the compositional meaning theory itself need not assert any pathological sentences. This undermines at least one motive for denying that there are any genuine pathologies in natural languages, namely, that unless we find natural languages to be free of pathologies, there would be no coherent way of explaining to ourselves how our language works. Section III is a brief summary.

I.

Concepts are common elements in different thought contents. For present purposes, we restrict concepts to the sort of thing that correlates in thought with predicate expressions, though having a concept does not require having a predicate that expresses it. For example, the introduction of a new color term, which marks a distinction not previously marked in the language, involves the introduction of a term to express an antecedently possessed concept. Possession of a concept we take to be constituted by a competence in its correct deployment in reasoning and in response both to what one believes about a situation and to experience of both inner and outer sense.

Concepts, as we understand them, mark sharp boundaries between kinds of things. We may allow concepts to be properly applicable to only a range of kinds of thing (e.g., color concepts we may want to say are applicable only to things with surfaces), but within their range the concept either applies or fails-to-apply to everything. On our use of the term, then, it does not make sense to speak of vague concepts or partial concepts or incoherent concepts, concepts competence in deployment of which involves commitment to, e.g., contradictions.[4]

The job of a purely qualitative predicate (henceforth "predicate") is to express a concept. If a predicate expresses a concept, we will say that it is *fully meaningful* and *has a meaning*. The application conditions of a fully meaningful predicate are determined by the concept it expresses. Not every predicate whose use we learn in a language is fully meaningful or expresses a concept. In an artificial language such as a predicate calculus, the predicate letters are uninterpreted. They do not express concepts and are not fully meaningful. Yet they have associated with them a semantic category, that of predicate, and so are not simply completely meaningless signs. Similarly, not every predicate in a natural language expresses a concept. Some nonsense predicates which are arguably a part of the language are like predicate letters in the predicate calculus, e.g., "mimsy" from Lewis Caroll's nonsense poem "Jabberwocky". More significantly, vague predicates do not express

concepts (hence, do not have meanings and are not fully meaningful) in the sense in which we use the term. However, it would be problematic to say that they are meaningless. They are learned when we learn a language, they are common currency in speech and communication, and their use is subject to normative constraints which impose a distinction between corrections on the basis of fact and on the basis of conventions of use. We will say that such terms do not have meanings, but they have a *meaning practice* associated with them that determines some *rules of use* for them in the language. An expression which has no meaning practices associated with it in a community is meaningless in that community's language. If it has a meaning practice associated with it, it may yet lack a meaning, though it serves its function well enough.

Learning a language is a matter of being trained in its use. When we speak in default contexts with literal intent, we speak with the intention of using our words in accordance with a community practice. The meaning practices, and, hence, what meanings words have, are determined by a complex of interlocking conditional joint intentions with respect to their use by members of the community. We will speak of these intentions as *meaning commitments* or *meaning intentions*. In the rare case in which a word is introduced explicitly into the language, the introducer may intend reflectively that it track a certain property or express a certain concept or obey certain principles. We will call this a meaning intention and a meaning commitment as well.

We distinguish between the case in which a term is used with the intention that it express a certain antecedently grasped concept, and the case in which a term is introduced by way of a set of principles that is supposed to govern its behavior (for example, a connective introduced by way of introduction and elimination rules). In the latter case, the concept it expresses, if any, is determined by the principles imposed on its use. We will call these "meaning principles".[5] In the former case, the meaning principles governing its use are determined by the concept it is intended to express, *provided that* the intention that it express it is successful. If the intention is not successful, we will say that there are no *meaning principles* governing the predicate, though there may yet be linguistic practices attaching to it and corresponding rules of use governing the practice ("rules" for short). If there are meaning principles governing a term, there are rules of use, but there may be rules of use without there being meaning principles. The relevance of these distinctions will become clear below.

It is an empirical matter what any given expression in a natural language means or what meaning practices are associated with it. We aim to give a plausibility argument for the truth predicate having incoherent meaning practices associated with it and failing to express the concept of truth, though it is intended to do so. This consists of two parts.

The first is the observation that natural languages are not social institutions that were introduced from a reflective standpoint but have rather evolved haphazardly under a variety of pressures and influences. They are redundant, ambiguous, vague and imprecise. They are far from perfect instruments of their kind. While sophisticated in many ways, there is no *a priori* or empirical presumption that natural languages are perfectly okay just as they are. There is no particular reason to suppose that the work-a-day practices associated with predicates must always result from a set of meaning commitments which yield a precise and consistent set of application conditions. It would be astonishing, in fact, if ordinary speakers managed to anticipate in their meaning commitments all the situations in which questions or problems might arise. We should expect otherwise when some subtle difficulty arises out of what would seem a natural development in the expressive powers of language.

The second is to provide a model of how the natural inclination to introduce a predicate to express the concept of truth runs into such a subtle difficulty and why it would be a mistake to suppose right off that the difficulty reflects on the concept of truth rather than on a limitation on its expressibility in a particular sort of symbolic representational medium it is about.

Consider an object language $\mathfrak{L}$ and a metalanguage $\mathfrak{M}$. We will indicate just enough about their syntax and structure to carry us through the discussion. $\mathfrak{L}$ contains the following connectives, "$\wedge$", "$\vee$", "$\neg$", "$\Rightarrow$", "$\Leftrightarrow$", which are interpreted classically as conjunction, disjunction, negation, the material conditional and biconditional respectively, while $\mathfrak{M}$ contains correspondingly, "&", "or", "$\sim$", "$\rightarrow$", "$\leftrightarrow$". Object language predicates are upper case Roman letters from "A" to "L" with or without subscripts and with superscripts indicating their arity. Lower case Greek letters "α" to "λ" with or without subscripts are object language metalinguistic names. "λ" refers to "$\neg T(\lambda)$", though this is not a sentence of $\mathfrak{L}$. We will use an Allegro typeface for $\mathfrak{M}$ to distinguish it from $\mathfrak{L}$. We will suppose that for every predicate of $\mathfrak{L}$ there is a predicate of $\mathfrak{M}$ which translates it. $\mathfrak{L}$ and $\mathfrak{M}$ are context insensitive and lack quantifiers. In our discussion of these languages our metalanguage is an extension of English which includes specialized semantic predicates which will cover appropriate extensions of $\mathfrak{L}$ and $\mathfrak{M}$ but not English or any other natural language, and generally speaking are so constrained as not themselves to lead to any untoward results. We will use underlining to indicate the use of such predicates. This will enable us to talk about the semantic properties of expressions in $\mathfrak{L}$ and $\mathfrak{M}$ without using suspect predicates.

Suppose that "*True(x)*" is a truth predicate in $\mathfrak{M}$ that expresses the concept of <u>truth</u> for $\mathfrak{L}$ and extensions of $\mathfrak{L}$, where an extension of $\mathfrak{L}$ preserves the vocabulary and interpretation of $\mathfrak{L}$ but contains additional predicates drawn from "M" to "Z" with or without subscripts and with superscripts

indicating arity. We will refer to this concept with "the concept of **truth**". It is clear that no difficulty arises in the application of "True(x)" to sentences of $\mathfrak{L}$. It is clear that the concept expressed by "**true(x)**" and so its meaning underwrites instances of the T-schema, (T), where "**s**" is replaced by a name or description of a *fully meaningful and non-defective sentence* of $\mathfrak{L}$ and "**p**" is replaced by a sentence of $\mathfrak{M}$ that translates it, that is to say, all such instances are <u>true</u>.

(T)   **true**(**s**)↔**p**

The schema will apply also to all extensions of $\mathfrak{L}$. However, if we introduce an extension of $\mathfrak{L}$ which contains a term that is not translatable into $\mathfrak{M}$, there will not be an appropriate instance of (T) in $\mathfrak{M}$ for sentences containing that term, though we could extend $\mathfrak{M}$ to accommodate one. The restriction of the schema to fully meaningful sentences is important. Although we stipulate that the basic vocabulary of $\mathfrak{L}$ is fully meaningful and non-defective, we will allow extensions of $\mathfrak{L}$ to contain nonsense terms, vague expressions, and other terms which perhaps display some pathological behavior. Instances of (T) with respect to such sentences are not underwritten by the concept of **truth**, even if there are translations of such sentences into $\mathfrak{M}$, so that there are instances of (T) in $\mathfrak{M}$ with respect to them. More generally, a sentence falls under the concept of **truth** or **falsity** only if it is fully meaningful and non-defective.[6]

Suppose we now wish to introduce an extension of $\mathfrak{L}$, $\mathfrak{L}^+$, in which the concept expressed by "**true(x)**" is expressed by some predicate "T(x)" of $\mathfrak{L}^+$. That is, our primary meaning intention is to use "T(x)" to express in $\mathfrak{L}^+$ the concept expressed by "**true(x)**" in $\mathfrak{M}$. Let us assume that this is possible. Then "T(x)" in $\mathfrak{L}^+$ translates "**true(x)**" in $\mathfrak{M}$. Thus, if the <u>truth</u> of a sentence in $\mathfrak{M}$ is underwritten by the concept expressed in $\mathfrak{M}$ by "True(x)", then any sentence of $\mathfrak{L}^+$ which is a translation of it will be underwritten by the concept expressed by "T(x)" in $\mathfrak{L}^+$, that is, it will likewise be <u>true</u>. Thus, if "T(x)" is to express the concept of "**true(x)**" in $\mathfrak{L}^+$, its meaning must underwrite all instances of (1) which are translations of instances of (T) in $\mathfrak{M}$ into $\mathfrak{L}^+$, where "p" is replaced by the sentence denoted by "s". Given that "λ" in $\mathfrak{L}$ denotes "¬T(λ)", (2) is an instance of (1) underwritten by the meaning of "T(x)", given our other assumptions. So, if "T(x)" expresses the concept of **truth** as intended, (2) is <u>true</u>, and, hence, in the extension of "**true(x)**" in $\mathfrak{M}$.

(1)   T(s) ⇔ p
(2)   T(λ) ⇔ ¬T(λ)

However, the *meaning principles* already associated with the connectives in $\mathfrak{L}^+$, inherited from $\mathfrak{L}$, require that (2) does not fall in the extension of "**true(x)**" in $\mathfrak{M}$, and hence that it does not fall under the concept of **truth** for $\mathfrak{L}^+$.[7] This shows that the attempt to extend $\mathfrak{L}$ by a predicate that expresses the

concept expressed by "$\boldsymbol{true(x)}$" fails. More specifically (assuming relativiza-
tion of "T(x)" to $\mathfrak{L}^+$ and "$\boldsymbol{true(x)}$" to $\mathfrak{M}$ throughout):

(i)   "T(x)" expresses the same concept as "$\boldsymbol{true(x)}$", namely, the concept
        of **truth**, only if "T(x)" translates "$\boldsymbol{true(x)}$".
(ii)   If "T(x)" translates "$\boldsymbol{true(x)}$", then instances of (1) in which "p" is
        synonymous with s (relevant instances of (1)) translate instances of
        (T) in which "$\boldsymbol{s}$" translates "s" and "$\boldsymbol{p}$" translates "p" (corresponding
        instances of (T)).
(iii)   If relevant instances of (1) translate corresponding instances of (T),
        then relevant instances of (1) fall under the concept of **truth**.
(iv)   If relevant instances of (1) fall under the concept of **truth**, then (2) falls
        under the concept of **truth**.
(v)   (2) does not fall under the concept of **truth**.
(vi)   "T(x)" does not translate "$\boldsymbol{true(x)}$". [(ii)-(v)]
(vii)   "T(x)" does not express the same concept as "$\boldsymbol{true(x)}$", i.e., "T(x)"
        does not express the concept of **truth**. [(i), (vi)]

There was nothing special about our choice of "T(x)", so the conclusion
generalizes: no predicate of an extension of $\mathfrak{L}$ can translate "$\boldsymbol{true(x)}$" or
express the concept of **truth**.

    Our primary meaning intention with respect to "T(x)" was that it express
the concept of **truth**. This led to the instrumental intention that it be
governed by (1), as this is a necessary, though not a sufficient, condition for
"T(x)" to express the concept of **truth**. Its being governed by (1), however,
precludes its expressing the concept of **truth**. Thus, our primary intention
cannot be carried out, and this shows that just having the intention that
"T(x)" express the concept of **truth** is not sufficient for it to do so.

    If our meaning intentions with respect to "T(x)" are unsuccessful, then we
fail to invest "T(x)" with a meaning.[8] For "T(x)" has a meaning only if it
expresses a concept. But given our intention, if it expresses any concept, it
expresses the concept of **truth**. It does not. So it expresses no concept, and,
hence, fails to have a meaning. This then tells us what the status is of the
sentences of $\mathfrak{L}^+$ containing "T(x)", namely, that they do not fall under either
the concept of **truth** or of **falsity**. Rather, as in the case of nonsense
sentences, they satisfy "$\sim \boldsymbol{true(x)}$". Consequently, nothing follows from
instances of (1).

    In light of this, what should we say about the semantic status of "T(x)"?
We earlier distinguished between intending that an expression have
whatever meaning is induced by its obeying certain principles and intending
that an expression express a concept, so that the meaning principles
governing it derive from that requirement. In the case of "T(x)", the
meaning intention was of the latter sort. Contrast this with the introduction
of a predicate "T*(x)" where the intent is not that it express the antecedently
grasped concept of **truth** but that its meaning sanction relevant instances of

(1).[9] In this latter case, it is clear that the meaning principles introduced are inconsistent in the sense that adhering to them requires sanctioning a contradiction. Here we may speak of the predicate being semantically inconsistent.

But it is different in the case of "T(x)". In the case of "T*(x)", our meaning intentions are successfully carried out and they attach to the predicate inconsistent meaning principles. In the case of "T(x)", however, our meaning intentions are not successful. "T(x)" cannot express the concept it is intended to. It would be governed by meaning principles dictated by the concept of **truth**, if it could express it. But it cannot. Therefore, no meaning principles are imposed on "T(x)". It is not then semantically inconsistent in the sense in which "T*(x)" is.

What do we say in the case in which we fail to notice that the attempt to let (1) govern "T(x)" leads to a formal contradiction? In this case, we proceed as if we were successful and so there will be meaning practices and corresponding rules associated with it. That practice is incoherent, and the associated rules are inconsistent. But this is not semantic inconsistency, because we are under the illusion that our semantic intentions were successful when they were not.

It is sometimes suggested that understanding truth consists in a disposition to accept relevant instances of the T-schema. Perhaps then it will be thought that after all the intended meaning of "T(x)" is properly and fully characterized by the relevant T-biconditionals and that the result, (2), rather than showing that "**True(x)**" is not translated by "T(x)", shows that it too is defective. We can show that this thought is incorrect. Suppose one adds two truth-predicates, "$T_1(x)$" and "$T_2(x)$", to an extension of $\mathfrak{L}$, $\mathfrak{L}^{2+}$, with the intention that they express the concept of **truth**, and suppose that λ_1 and λ_2 refer respectively to "$\neg T_1(\lambda_1)$" and "$\neg T_2(\lambda_2)$", and λ_3 and λ_4 refer respectively to "$\neg T_2(\lambda_1)$" and "$\neg T_1(\lambda_2)$". The two predicates should have the same meaning, because they are intended to capture the same conceptual content, and on the view in question the two corresponding T-schemas, (T1) and (T2), determine the same meaning.

(T1) $T_1(s) \Leftrightarrow p$

(T2) $T_2(s) \Leftrightarrow p$

Nevertheless, they do not. For λ_1 and λ_2, given (T1) and (T2), lead directly to contradictions and so are pathological, while the result of replacing each truth predicate in these sentences by the other, λ_3 and λ_4, do not, but rather seem to say the right thing about the pathological sentences λ_1 and λ_2. T_1 and T_2 do not have the same meaning then, and it follows that one or the other fails to express the concept of truth with respect to the language. Since there is perfect symmetry between them, the proper conclusion is that

neither does. Therefore, the relevant instances of (1) do not properly and fully characterize the intended meaning and their inconsistency is not properly characterized as semantic.

The purpose of this discussion has been to provide a model for a certain picture of what has gone wrong with semantic predicates in natural languages and a basis for understanding what the appropriate response is in various contexts. Our metalanguage $\mathfrak{M}$ is modeling the level of our thinking about language, and our object language $\mathfrak{L}$ and its various extensions model natural languages and their evolution. We are, it seems, able to think about the truth of sentences and utterances. This is a conceptual capacity which may in principle be exercised in the absence of having introduced a predicate in any language one speaks to express the concept deployed. If this is right, we can see why a difficulty arises when we try to introduce into a language a predicate which expresses the concept we can deploy in thought about it, and why that need not reflect back on the concept which we thereby attempt to express in the language to which we apply it. The concept of truth, we suggest, is not problematic, but it is rather the truth predicate, which we intend to express the concept of truth, which is problematic—and which fails to play the role intended for it.

It is easy to see, in terms of our model, how it can seem unproblematic to introduce a truth predicate into a language to express the concept of truth for it, despite there being an in principle difficulty, and it is also easy to see why we typically do not get into trouble in using the truth predicate. In the case of $\mathfrak{L}^+$, it is clear that what we intend to say in $\mathfrak{L}^+$ can be said in $\mathfrak{M}$, and it is clear that the thought on the basis of which we use "T(x)" in $\mathfrak{L}^+$ is one that need not itself be problematic in the way the sentence we use is. Thus, it is easy to see how communication can proceed despite the imperfections in the language we use to communicate our thoughts. This, we suggest, is the situation with respect to natural language truth predicates as well. Our meaning intention is to use "true" in English to express the concept of truth. What we think we fail to express in the language, for the reasons illustrated in our model, but this is not a barrier to communication, because the thoughts on the basis of which we use sentences containing the truth predicate need not be infected by its deficiencies. We see through the language, so to speak, straight to the thoughts it is intended to express because we know what meaning intentions are connected with the predicate, though the predicate fails to fulfill the role intended for it.[10]

We have used the truth predicate in the course of the foregoing discussion. On our own account, our meaning intentions with respect to it fail to secure for it a meaning, though they induce a meaning practice with respect to it, which has an appropriate context sensitive T-schema governing it. If we are right, this is no barrier to communication, for the thoughts which our words fail to express are nonetheless evident in virtue of the grasp English speakers have on the concept of truth and our meaning commitments with respect to

the truth predicate. (*Mutatis mutandis* for other expressions in natural languages whose understanding is bound up with the semantic predicates.) Our thesis is an empirical one, being about natural languages. But it is, we believe, antecedently plausible that we have a coherent concept of truth, and this is supported by the fact that we are able to devise ways of talking about truth and falsity in natural languages which do not generate the difficulties which our meaning practices with extant truth predicates appear to run into. These alternative ways of talking about truth and falsity designed to avoid the difficulties our native predicates are involved in are evidently guided by our understanding of what has gone wrong, what *cannot* be true, and how it is that we can sidestep the difficulties by various devices (context sensitive truth predicates, Kripke's construction, moving to a metalanguage to talk about truth in a given language). Given that we have a concept of truth, it would be natural to expect in the unscientific and haphazard development of natural languages that a predicate would be introduced to express it governed by a suitably general but therefore fatal T-schema and that the difficulties would remain unnoticed until reflection on natural languages had reached a fairly sophisticated level. It seems to us implausible to suppose that natural languages as practical social institutions would have developed a sophisti-cated set of semantic principles that anticipated a subtle difficulty in the enterprise of expressing the concept of truth when even the requisite tools for thinking carefully about them would not have been available. If this is right, then various proposals which have been made about the actual semantics of the truth predicate which are designed to show that the semantic principles attaching to it are sophisticated enough to avoid any difficulties are better thought of as instructions for how to revise the language self-consciously so as to avoid the difficulties into which we have fallen.[11]

II.

One motivation for finding natural languages, appearances to the contrary notwithstanding, to be alright just as they are is that admitting that semantic predicates have inconsistent meaning practices associated with them threatens to throw into confusion not only much ordinary talk but also our best theoretical framework for getting a handle on the semantics of natural languages. A semantics for a natural language may be thought of as dividing into two parts. One part is concerned to explain the meaning of primitive expressions in the language. Another is concerned to explain, granting the meanings of the primitive expressions, how we understand the meaningful complexes formed from them. In both cases, it is natural to try to explain expression meaning by what satisfaction or truth conditions it induces on the expression. But there appear to be significant difficulties in the way of carrying out this program in the case of expressions containing semantic predicates, if what we have urged is correct, for then semantic

predicates do not strictly speaking have application conditions, and sentences which use them do not have truth conditions. One response is to treat only those portions of the language which do not contain semantic predicates and to use a metalanguage with a restricted truth predicate that aims to cover just the language with its semantic terminology excised. But rather than settle for this we will argue that (i) even for defective predicates there is a surrogate for providing satisfaction conditions and that (ii) the purposes of a compositional meaning theory can be served by an axiomatic truth theory in application to a language containing defective predicates without their infecting the meaning theory itself.[12] We begin with (ii).

A compositional semantics in the Davidsonian tradition makes central use of an axiomatic truth theory for a language.[13] For purposes of illustration, let us use $\mathfrak{L}$ and $\mathfrak{M}$ from the previous section, where $\mathfrak{M}$ is the language of the theory and $\mathfrak{L}$ is the language for which the theory is given. The truth theory does its work by producing for each object language sentence, from axioms using metalanguage expressions which interpret the object language expressions for which they give satisfaction conditions (interpretive axioms), a theorem of the form (T), repeated here, where ''**s**'' is replaced by a structural description of a sentence of $\mathfrak{L}$ and ''**p**'' is replaced by a sentence of $\mathfrak{M}$ that translates it.

(T)   𝔗rue(**s**)↔**p**

If the theory has interpretive axioms, we will call it interpretive. A theorem that is an instance of (T) meeting the conditions noted we will call an *interpretive T-theorem*, or a *T-sentence* for short. Intuitively, a proof which draws solely on the content of interpretive axioms in proving a theorem of the form (T) will be interpretive. We will call such a proof procedure a *canonical proof procedure*, and a theorem of the form (T) arrived at via such a procedure a *canonical theorem*. If we have an interpretive truth theory and a canonical proof procedure, we are in a position to produce a T-sentence for each sentence of the object language via a proof which reveals at each stage how the immediate constituents of a complex expression contribute to fixing its truth conditions by way of axioms which exhibit what they mean. Thus, the interpretive truth theory plus appropriate knowledge about it puts us in a position to interpret any object language sentences by way of a procedure which exhibits how its constituents contribute to determining its interpretive truth conditions. Once we have a canonical theorem, knowing that ''**p**'' translates **s**, we can infer from the appropriate instances of (T) the corresponding instances of (M).

(M)   **s means that p**

So far so good.

Consider, however, the application of this program to a language containing a predicate which is intended to express the concept of truth with respect to the language containing it, in the manner of $\mathfrak{L}^+$. In our terminology, the predicate "T(x)" in $\mathfrak{L}^+$ does not have a meaning (it expresses no concept in $\mathfrak{L}^+$), but it is not meaningless because it has a meaning practice associated with it (imagine the difficulty to be undiscovered). If we apply this approach to a language in which not all the expressions are fully meaningful, then we must treat "translates" and "means" as relating expressions on the basis of sameness of associated meaning practices. This allows it to relate expressions which, though not meaningless because they have a meaning practice associated with them, are not fully meaningful and so don't have satisfaction or truth conditions.[14] The goal of a compositional meaning theory in the style just illustrated would be to produce an interpretive T-theorem and so a corresponding theorem of the form (M) as an explicit statement of the meaning of object language sentences containing "T(x)". This requires that we have a predicate in our metalanguage which translates "T(x)". It is not "**true(x)**" since we have proved that "T(x)" does not express the concept of **truth**. Suppose that "$\boldsymbol{\tau(x)}$" is such a predicate, and that "**l**" in $\mathfrak{M}$ translates "λ" in $\mathfrak{L}^+$. Then the canonical T-theorem for λ would be (5).

(5) **true("**$\neg$T(λ)**")** $\leftrightarrow$ $\sim\boldsymbol{\tau(l)}$

Our difficulty is not that this is itself a contradiction or that it reduces to one. If "$\boldsymbol{\tau(x)}$" and "**true(x)**" had the same extension, then (6) would be a material consequence.

(6) $\boldsymbol{\tau(l)} \leftrightarrow \sim\boldsymbol{\tau(l)}$

But they do not. "T(x)" cannot have even the extension of "**true(x)**" on pain of contradiction, so likewise its translation "$\boldsymbol{\tau(x)}$" does not have the extension of "**true(x)**". The difficulty is that (5) is not <u>true</u> (or <u>false</u>). If our axiomatic truth theory is a part of our compositional meaning theory, then it seems that to carry out the program for a language with defective predicates, like truth predicates in natural languages, we must endorse a theory that is not true because the axioms for the defective predicates are not true. But surely any correct compositional meaning theory for a language must be true.

The solution to the difficulty is to recognize that insofar as the meaning theory is constituted by what we have to know about a truth theory to use it to specify the meanings of any sentences in the language in a way that exhibits how the parts contribute to the meanings of the whole, the meaning theory does not involve what the truth theory says as opposed to propositions about the truth theory. More specifically, what we need to know about a truth theory $\mathfrak{T}$ as indicated above is of the following form:

(i)   From a canonical theorem of an interpretive truth theory $\mathfrak{T}$ for a language L the corresponding M-sentence follows;

(ii) $\mathfrak{T}$ is an interpretive truth theory for L whose axioms are …;

(iii)   Axiom … of $\mathfrak{T}$ means that …;
        axiom … of $\mathfrak{T}$ means that …;

   …

(iv)   … is a canonical proof procedure for $\mathfrak{T}$.

In (i), an M-sentence will be a sentence in the language of the theory that translates an appropriate instance of "s means that p". In a filling in of this schema, (ii)-(iv) state what we need to know about a truth theory to derive and understand a canonical theorem for any declarative sentence of the language, while (i) states what we know about the connection between being a canonical theorem and a corresponding statement about what a sentence means, which enables us to infer from each canonical theorem a sentence we understand which states what the object language sentence means. The language of our explicit meaning theory need not be the same as that for our truth theory, and the axioms of the truth theory are not axioms of the meaning theory. Thus, even if some axioms of the truth theory are not true, that does not affect the truth of the meaning theory. The pathology of a predicate which would undercut the possibility of a sentence in which it would need to contribute its extensional properties to the sentence's truth conditions does not undercut the truth of an M-sentence when it appears on the right hand side because the M-sentence is true provided that the sentence in which the predicate appears in the complement translates the sentence mentioned by the subject term. Consider $\mathfrak{L}^+$ again. The M-sentence for λ in $\mathfrak{L}^+$, as stated in, say, $\mathfrak{M}$, is (7).

(7)   "¬T(λ)" **means that** ~𝛕(l)

(7) is <u>true</u> just in case "~𝛕(l)" in $\mathfrak{M}$ translates "¬T(λ)" in $\mathfrak{L}$.[15] It does, and so (7) is <u>true</u>, though neither "~$\mathfrak{T}$(l)" nor "¬T(λ)" are. The truth theory though still serves its dual purpose of enabling us to produce a meaning specification of each sentence of the object language and in a way that shows how each expression is supposed to contribute to what it means by articulating its contribution to a canonical theorem in a truth theory for the language. If the axioms involved in a proof are not problematic, then the proof shows how each expression contributes to fixing the interpretive truth conditions for the sentence. If one or more axioms in a proof involve semantically defective expressions, then we see at least how the expression is supposed to contribute, so far as its logical form and category go, to the schematic truth conditions for expressions containing it, and so how compositionally it contributes what it does to the meaning (or meaning practices associated with) the sentences in which they appear. Allowing

relevant instances of $\mathfrak{M}$ to be true even when the sentence it is about does not have truth conditions does not mean that we have broken the connection between meaning and truth. Rather, the connection is this: if s means that p and s is semantically complete and coherent, then s is true iff p. A statement of truth conditions for each semantically complete and coherent sentence of the object language then does follow from the meaning theory, though to know in which cases the inference is sound we need to know which predicates in the object language are defective.

When we turn to the question of lexical semantics for semantically defective predicates, we may of course produce the analog of an M-sentence for such a predicate, for example, (8).

(8) "T(x)" *means τ(x)*.

This is not without point when the language is not ours, but for our own language it does not advance us beyond where we started. What we want in the case of a semantically complete predicate is an account of what is involved in understanding the expression. This may take the form of a conceptual analysis, but, as not all terms will admit of analyses in more basic terms, in some cases this will take the form of an account of the possession conditions of the concept where this involves relations to other concepts and to application conditions specified in terms perhaps of perceptual experience relative to various background beliefs. However, none of this will be available in the case of a term which does not express a concept, as in the case of the truth predicate in natural languages. What we can advert to in this case is not what the term means but what our meaning commitments are with respect to it, and how they induce on the predicate the meaning practices that attach to it. In this respect, we believe that the best that can be done is to undertake the kind of diagnostic account sketched in section I of this paper.

III.

We have aimed to make plausible the view that the semantic *concepts* for natural language are unproblematic and that what is problematic is the prima facie unexceptionable intention to express them in the languages to which they apply. It cannot be done, and truth predicates in natural languages fail to express the concept of truth, or any other concept. If this is right, then there are things we can think about the languages we speak which we cannot express in them (*mutatis mutandis* for other semantic concepts). This means that natural languages are not universal in the sense of being able to express their own semantic concepts. The assumption that they are universal in this sense together with the pathologies exhibited by the semantic predicates leads to the view that our semantic concepts are likewise

pathological. We hope to have shown that there is no good reason for this assumption. The view that the concept of truth is fine but that the truth predicate is not helps to explain why we can say sensible things about what has gone wrong and also how we manage to communicate about truth using predicates which fail to express the concept: for we do use the predicates with the intention that they express a concept which we antecedently grasp, and we all know this. We have suggested that we need a more nuanced view of what has gone wrong with the semantic predicates than that they are semantically inconsistent, this being reserved for terms for which our meaning intentions have been successful but upon which they have imposed inconsistent principles. As our meaning intentions with respect to the semantic predicates are unsuccessful, we have not attached meaning principles to them. But being under the illusion they are successful, a practice in their use arises which is inconsistent in the sense that the associated rules of use are inconsistent. We have also argued that admitting that the semantic predicates of natural languages are in this way problematic does not, as might have been thought, undercut the possibility of giving a compositional semantics for the language by appeal to an interpretive truth theory for it. The key insight is that the truth theory is not itself the meaning theory. Rather, information about the truth theory is exploited for the purposes of a meaning theory. This means that the truth theory need not be true to serve its purposes and that the meaning theory itself need not assert any sentence in which a problematic semantic predicate is used in a way that requires its extensional properties to contribute to fixing the sentence's truth conditions. In the case of lexical semantics, we urge in the case of pathological predicates an account of the meaning intentions on the basis of which they are used and how and why they fail to succeed in giving a predicate the role of expressing a concept and how and why they induce on the term the meaning practice they do, despite this fact.

Notes

1. For example, Barwise and Etchemendy (1987); Berk (2004); Glanzberg (2004); Greenough (2001); Grover (1977); Simmons (1993) among others.
2. This is the approach which Graham Priest and his followers have adopted, see Priest (2006).
3. We thus give up the assumption that natural languages are universal in the sense of being capable of expressing their own semantic concepts. This contrasts with many positions in the literature where the assumption is that *since* natural languages are *universal* (with respect to their capacity to express their own semantic concepts), their semantic concepts are themselves *inconsistent*. For example, Azzouni (2003); Patterson (2005); Priest (2006).
4. We follow the Fregean tradition here, using ''concept'' in the sense of Fregean *Sinn*. We believe this lends clarity to the discussion of the truth predicate's relation to our commitments in using it. ''Concept'' is also often used as a correlate of the dispositions acquired by inculcation into a community's practice in the use of a word. This usage

blurs the distinction between our thoughts and intentions in using language and the practices that they induce on the words we use.

5. Whether this lines up with Eklund's term "meaning constitutive principle" depends in part on how one wants to understand "semantic competence", in terms of which he introduces it: "a principle is constitutive of the meaning of an expression just in case competence with the expression involves being disposed to accept the principle as true or valid" (2002, p. 256). Anything we call a "meaning principle" will be a meaning constitutive principle in Eklund's sense. However, we mean to draw a distinction between meaning principles and rules of use, where the latter are induced by a meaning practice with a term when there may be no meaning principles associated with it. If being inculcated into such a practice counts as semantic competence in the use of the term, then our use of "meaning principle" is narrower than Eklund's. If "semantic competence" can be had only with terms for which our meaning intentions are successful, then our use of "meaning principle" will line up with Eklund's use of "meaning constitutive principle".

6. This may seem too stringent in the case of vague predicates, if vagueness is construed as a sort of semantic incompleteness. But if there is higher order vagueness there is nonetheless a sound argument to show that no sentence using a vague predicate is true or false. See Ludwig and Ray (2002).

7. Extending $\mathfrak{L}$ with a new predicate does not change the meanings of any of the terms already in $\mathfrak{L}$, so the connectives are still to be construed classically. For the same reason, the truth conditions of sentences that $\mathfrak{L}$ and $\mathfrak{L}^+$ share will be the same and will be alike in truth value.

8. We thus do not accept the assumption that Patterson suggests is behind much orthodox work on the semantic paradoxes, and which he incorrectly attributes to Ludwig in Patterson (forthcoming), namely, that speakers cannot be wrong about whether their meaning intentions are successful.

9. Chihara (1979) distinguishes in the same way between "true" and "true*", but he takes both predicates to be semantically inconsistent.

10. It will be natural to worry that the pathologies that arise with respect to the meaning practices associated with semantic predicates in natural languages will arise again at the level of thought. However, some unpublished work by Badici on the relation between the functional role signified by psychological mode and attitude content provides a model for an account of the consistent deployment of the concept of truth in thought.

11. Examples of these proposals include Simmons (1993), Kripke (1975) and McGee (1990).

12. The following discussion is prefigured in Lepore and Ludwig (2005, chapters 9-10) and Ludwig (2002).

13. For the record: we do not subscribe, and neither did Davidson, to the view that semantic competence consists in propositional knowledge of a truth theory or any other theory of the language. The goal of a meaning theory is rather, in Dummett's words, to provide "a theoretical representation of a practical ability", that is, the ability to speak and understand a language.

14. Patterson (forthcoming) takes Ludwig in earlier work to be committed to interpretive axioms genuinely assigning satisfaction or truth conditions, and so to be true, even for the semantic predicates. This is not Ludwig's view. Interpretiveness in the case of languages containing pathological or semantically incomplete expressions comes to no more than that the relevant expressions used on the left hand sides of our axioms have the same meaning practices associated with them as the expressions mentioned on the right.

15. We adopt a sententialist construal of "x means that p", so that "that p" is taken to refer to "p"; in taking "x means that p" to be true just in case x translates "p", then, we capture both cases in which "p", as it would be said, expresses a proposition (is fully

meaningful) and cases in which it does not yet still has a meaning practice associated with it. Otherwise no meaning theory of this form can be given for a natural language some of whose declarative sentences, because of semantic defects of contained expressions, do not express propositions. (Alternatively, one could introduce entities to associate with sentences which are the same when the associated meaning practices are.)

References

Azzouni, J. (2003) ''The strengthened liar, the expressive strength of natural languages, and regimentation'', *Philosophical Forum*, 34, pp. 329–350.

Barwise, J. & Etchemendy, J. (1987) *The Liar: an Essay in Truth and Circularity* (Oxford: Oxford University Press).

Berk, L. (2004) ''The Liar, context and logical form'', *Journal of Logic, Language and Information*, 13, pp. 267–286.

Chihara, C. (1979) ''The semantic paradoxes: a diagnostic investigation'', *The Philosophical Review*, 88, pp. 590–618.

Eklund, M. (2002) ''Inconsistent languages'', *Philosophy and Phenomenological Research*, 64, pp. 251–275.

Glanzberg, M. (2004) ''A contextual-hierarchical approach to truth and the liar paradox'', *Journal of Philosophical Logic*, 33, pp. 27–88.

Greenough, P. (2001) ''Free assumptions and the liar paradox'', *American Philosophical Quarterly*, 38, pp. 115–135.

Grover, D. (1977) ''Inheritors and paradox'', *Journal of Philosophy*, 74, pp. 590–604.

Kripke, S. (1975) ''Outline of a theory of truth'', *Journal of Philosophy*, 72, pp. 690–716.

Lepore, E. & Ludwig, K. (2005) *Donald Davidson: Truth, Meaning, Language and Reality* (New York: Oxford University Press).

Ludwig, K. (2002) ''What is the Role of a Truth Theory in a Meaning Theory?'', in: D. Shier, J. K. Campbell & M. O'Rourke (Eds), *Meaning and Truth: Investigations in Philosophical Semantics*, pp. 142–163 (New York: Seven Bridges Press).

Ludwig, K. & Ray, G. (2002) ''Vagueness and the sorites paradox'', *Philosophical Perspectives*, 16, pp. 419–61.

McGee, V. (1990) *Truth, Vagueness, and Paradox* (Indianapolis: Hackett Publishing Company).

Patterson, D. (2005) ''Sentential truth, denominalization, and the Liar: aspects of the modest account of truth'', *Dialogue: Canadian Philosophical Review*, 44, pp. 527–37.

Patterson, D. (forthcoming) ''Inconsistency theories of semantic paradox'', *Philosophy and Phenomenological Research*.

Priest, G. (2006) *In Contradiction: A Study of the Transconsistent* (Oxford: Clarendon Press).

Simmons, K. (1993) *Universality and the Liar: An Essay on Truth and the Diagonal Argument* (New York: Cambridge University Press).