<u>Words by Convention</u>

Gail Leckie and Robert Williams[1]

A long-established project in the philosophy of language is the search for a reductive naturalistic metasemantics. Reductive metasemantics ground target semantic facts - such as the fact that the word type "cat" refers to cats - in a non-semantic base.[2] Existing reductive projects presuppose that word- (or sentence-) types like "cat" - are part of the non-semantic base. They presuppose the availability of an *exogenous* theory of word types, that is, one that is prior to and independent of the metasemantics.[3] This paper argues that an exogenous account of word types is unlikely to succeed. We propose a new strategy: an *endogenous* account of word types, that is, one where word types are fixed as part of the metasemantics. In particular, we show how a metasemantic account on the lines of Lewis' account in terms of conventions of truthfulness and trust can provide an endogenous account of words suited to a naturalistic metasemantics. We say that it is the conventions of truthfulness and trust that ground not only the meaning of the words (meaning by convention) but also what the word type is of each particular token utterance (words by convention).

We begin by explaining why existing exogenous theories of word types ill-serve the metasemantic project (section 1). We then detail how a Lewisian metasemantics in terms of conventions of trust and truthfulness would use an exogenous theory of words (section 2), before adapting that metasemantics to give an endogenous theory of words in terms of conventions (section 3). In section 4, we show that our endogenous account deals well with the problem cases we earlier identified for exogenous accounts of words. Finally we raise the potential problem of overgeneration of words for our account (section 5) and show that Lewis' account of convention already has the resources to prevent such overgeneration (section 6).

1.

[2] We use quote marks variously to pick out word and sentence types, orthographic types and particular token utterances. We hope it is clear which use is in play on a particular occasion.
[3] For example, Lewis (1983), Evans (1973), Davidson (1973), Horwich (1998). By contrast, Millikan (1984) and Richard (1990) may intend something more like what we call an endogenous theory of word types.

As an example of a metasemantic theory that requires an exogenous account of words, consider a simple version of a dominant causal source account, modelled on Evans (1973). Where u is a token utterance, x is an object or property, y is a word-type,

(Toy Evans) u refers to x iff u is of type y and x is the dominant causal source of tokenings of the type y.

Notice that (Toy Evans), like most metasemantic theories, specifies how features of the word *type* fix a word *token's* semantic features.[4] To say what the meaning is of a specific token, u, we need to know its type. For example, Peta's utterance on 1/1/1850 "I've visited Madagascar" refers to whatever is the dominant causal source of tokens of that type "Madagascar" which Peta tokened. Evans says that the dominant causal source of the word-type "Madagascar" is the island and so Peta's token refers to that island, not the mainland.

How are we to use a metasemantic theory, such as (Toy Evans), in our naturalistic project? We must establish that all the terms on the right-hand side of the biconditional are legitimate parts of the nonsemantic naturalistic base. It is fair to assume that the island is a legitimate part of the base, as is the noise which is Peta's utterance. What about word types? Are these a legitimate part of the non-semantic base? To show that they are, an advocate of (Toy Evans) would need a naturalistic account of word types; and likewise for other metasemantic theories that give the semantics of tokens in terms of their word type. There are two extant naturalistic accounts of word types – phonographemic and Kaplanian.[5] We show that both generate counterintuitive predictions about what utterances are of the same type. This matters because those counterintuitive typings look likely to deliver undesirable semantics once plugged into any particular metasemantics.

The default picture of words in the philosophical literature is based on phonetic and

---

[4] We lack space to consider theories which provide metasemantic accounts for individual tokens directly without appealing to word types.

[5] The metasemantic theories we are considering require a metaphysics of public words since their target is shared languages, not idiolects. For this reason we do not consider *lexemes*, construed as the individual's typing on expressions.

orthographic similarity.[6] Where *u* and *v* are token utterances,

> (Phonographemic) *u* and *v* are tokens of the same word iff they are spelt or pronounced the same.

Words are maximal equivalence classes of tokens in the same-word relation.[7] So the word type "Madagascar" might be the equivalence class of all tokens that are spelt M-A-D-A-G-A-S-C-A-R etc, or all tokens sounded /ˌmadəˈgaskə/. Let's call this the *phonographemic conception*.[8]

The phonographemic account types word counterintuitively. Sometimes, it is too fine-grained, typing separately what are intuitively tokens of the same word type. At other times, it is too coarse-grained, typing together what are intuitively tokens of different word. In itself, counterintuitiveness is not strong evidence against the phonographemic account. It would be if we were performing descriptive conceptual analysis on the folk concept word. But there is no reason to take folk intuitions about the extension of "word" as revealing the extension of the natural kind that is of significance to semantics and metasemantics. Rather, counterintuitive typings are problematic in so far as, once plugged into particular metasemantic accounts, they deliver crazy semantic assignments.

Where an exogenous theory of words is intuitively too fine-grained, it provides an impoverished input for the metasemantic theory. It misses relevant tokens from the word type of an utterance *u*. Accordingly, the metasemantic theory the exogenous theory feeds into misses facts relevant to the meaning of utterances of *u*. If the input is sufficiently impoverished, the metasemantic theory has too little to go on and delivers no determinate meaning for *u*. If the input is selectively impoverished, the metasemantics may assign *u* a meaning but one that is incompatible with any plausible semantic theory.

On the other hand, exogenous theories that are too coarse-grained contaminate the input

---

[6] For example, Stebbings (1935); Davidson (1979) p. 90; Haack, (1978) p. 75. Cappelen 1999 and Cappelen and Dever 2001 propose a more sophisticated version of the phonographemic approach.

[7] That is, tokens of a word are all same-word related and there are no ways to merge these equivalence classes to produce larger equivalence classes whose members are all same word related.

[8] Hawthorne and LePore (2011) call this *the form-theoretic conception*. It is part of the view that Kaplan (1990) critiques under the name 'the orthographic conception.'

to the metasemantic theory.   When the metasemantics determines the meaning of *u*, it contaminates facts that do bear on the meaning of *u* - those about other tokens of intuitively the same type as *u* - with facts that don't – those about tokens of an intuitively separate word type(s). This may render the meaning of *u* indeterminate or skew its meaning towards what is intuitively the meaning of the other 'word type'.

Consider some examples. There are two ways that the phonographemic conception cuts too fine and impoverishes the input to the metasemantic theory. (Phonographemic) does not group any distinct sounds and spelling types together. Yet, sometimes an intuitively unitary word type has instances with various spellings ("realize"/"realise") or various pronunciations ("tomato"). The phonographemic account also counterintuitively counts instances of M-A-D-A-G-A-S-C-A-R and of /ˌmadəˈɡaskə/ as two distinct 'words'. This opens up the risk that the various phonemes and graphemes are assigned different meanings by the metasemantics and the risk that the input to the metasemantic theory for any grapheme or phoneme is so impoverished as to fail to generate the correct meaning for tokens of it. Suppose, for example, that what we would ordinarily think of the word "okapi" is uttered once and only once in the distinctive North Welsh accent. The class of tokens phonetically exactly similar to that utterance may be a singleton. Further suppose that the single North Welsh token was uttered in front of a muddy zebra's backside, not an actual okapi. Plugging the phonographemic account of names into (Toy Evans) now delivers crazy results. (Toy Evans) directs us to the dominant causal source of the word type and since the type is a singleton, it directs to the causal source of that particular token – the zebra.

(Phonographemic) also cuts more coarsely than the typing that gives the desired input into the metasemantic theory. For example, far-flung tokens of the grapheme C-O-W, traced out by by aliens on Mars, ought not to determine the semantics of Earth tokens of that grapheme C-O-W. That would contaminate the input to the metasemantics.   Even within a language, tokens of a single grapheme don't always have the same meanings - consider lexical ambiguity. For example, tokens spelt B-A-N-K sometimes bear information about financial institutions and sometimes about the edges of rivers. The watery tokens ought not to determine the semantics of the financial tokens. Once more, (Toy Evans) will struggle to deliver the right semantic results for such 'words' since the counterintuitive typing skews what is the dominant causal source of the type.   This typing might leave it indeterminate whether a token of B-A-N-K is the financial kind or the watery kind. In other cases, the

causal source of the additional tokens will outweigh the dominant causal source of local tokens of that phonographeme.

Going disjunctive might mitigate this concern. Perhaps we could say that a disjunctive property can be the dominant causal source of a word. We might try to say that B-A-N-K has a disjunctive property as its extension: is-a-riverbank-or-a-financial bank. However, these referential assignments are likely to cause trouble for the semantics of whole sentences once combined with a compositional semantics. We do not want "There are three banks" to come out as true in cases where there are at most two river banks and one financial bank.[9]

It is tempting to resolve the problem of coarse-grainedness by distinguishing two words with the same spelling and pronunciation but which are different words in virtue of their different meanings.[10]

> (Plus Semantic) u and v are tokens of the same word iff they have the same meaning and the same phonographemic features.

Again, words are maximal classes of tokens in the same word relation. (Plus Semantic) rules that there are two homonyms spelt B-A-N-K and so we might hope to get a different dominant causal source of each by (Toy Evans). However, (Toy Evans) can't be combined with (Plus Semantic) on pain of circularity. The account of word types can't appeal to notions which the metasemantic theory is supposed to ground naturalistically.

The phonographemic account of words is not the only one on offer. Kaplan offers a different exogenous account of word type. He takes two word tokens to be of the same type iff they are in an appropriate causal relation to one another. The paradigm appropriate causal relation is intended repetition. Not all tokens of B-A-N-K need be tokens of the same word, provided they fail to be related by chains of intended repetition. Conversely, other pairs of

---

[9]Going disjunctive brings other problems for the Evansian. Some tokens of the intuitive word "cow" have their source in horses, and yet we do not want any such token to mean horse. To avoid the horse/cow problem, it is tempting to build in a condition that disadvantages disjunctions as dominant causal sources.

[10] Many text books, such as Larson and Segal (1995) treat ambiguity as a syntactic phenomenon so that "bank" corresponds to two words. Plausibly, this move requires words to be individuated partially in semantic terms.

tokens that are spelt or pronounced differently - for example, "realize" and "realise" – are in the right sort of relation.

Since Kaplan's account has never been spelt out in detail, it's difficult to assess. However, it, too, types words counterintuitively in ways that are likely to cause trouble for any metasemantic theories reliant on it. Kaplan's account cuts too coarsely in cases like the following. Suppose Xena produces a token /treɪn/ in attempting to convey that she is a train-engineer, which Zara mishears and then repeats as /dreɪn/, forming and passing on the belief that Xena is a drain-engineer. Xena and Zara's tokens /treɪn/ and /dreɪn/ would count by Kaplan's lights as tokens of the same word since Zara intentionally repeats Xena's utterance. Zara's friends and relations go on to repeat her utterance in the context of discussions of plumbing. Such tokens count as the same word as Zara's /dreɪn/ tokening, but also by transitivity, as the same word as Xena's /treɪn/ tokening. The danger is that we lose the obvious fact that there are two word-types in play here. *Inter alia*, this plays havoc in identifying a single dominant causal source, as required by (Toy Evans).[11] As with the phonographemic account, it is tempting to get the intuitively right word typing by adding a semantic element to the account. We want to count only those intentional-repetition-links that preserve meaning, in order to identify a tradition of usages focused on conveying information about a single subject-matter. But a reductive metasemantics can't, on pain of circularity, make use of an exogenous account of words that contains semantic elements.

What have we shown so far? Phonographemic and Kaplanian views attempt to offer an account of word types that is separate from, and independent of, the main metasemantic account. We have established problems with combining these views of the metaphysics of words with (Toy Evans). At this point, one project is to retain the architecture: metasemantics underpinned by an exogenous theory of word types, and to show how the difficulties identified above can be resolved by more sophisticated development of the component theories. However, we think that it is a mistake to assume that the overall theory will have this architecture. What follows illustrates an alternative, on which word-typing and semantic facts are settled simultaneously and the word-types are endogenous to the

---

[11] Another type of case where Kaplan's account cuts more coarsely than an intuitive typing of words is the case of substantial but gradual phonetic, graphemic and semantic change, including cases with fission structure but we lack space to discuss such cases.

metasemantics. Our illustration of an endogenous theory will use a Lewisian metasemantics, though the endogenous strategy could potentially be deployed in other frameworks.[12]

2.

In this section, we describe the metasemantic theory – Lewis' – that we will be adapting into an endogenous account (in section 3). We start by recapping Lewis' own metasemantic story in terms of conventions of truth and trust. Lewis himself seems to be working with an exogenous phonographemic picture of word types. He says language assigns meaning to "certain strings of types of sounds or marks" (Lewis 1983 p163). The exogenous theories already canvassed turn out to cause problems for Lewis similar to the ones afflicting (Toy Evans).

In "Language and Languages" Lewis says that languages, in the sense of abstract semantic theories, are functions from sentences to meanings. [13] Grammars are, *inter alia*, functions from the public lexicon (a generalisation of the category word types) to meanings which recursively generate the functions from sentences to meanings.   The job of a metasemantic account is to lay down conditions, in an illuminating, non-circular way, as to which of these abstract grammars and languages is in use by a population.

Lewis' metasemantic theory says that a certain language L is correct for that population if the conventions of truthfulness and trust for L prevail for that population P. A convention of truthfulness for a sentence type s, connecting it to the proposition p, is a regularity in usage—that members of the population utter tokens of s only if they believe that p—where this regularity is entrenched in the beliefs and desires of the community in a distinctive convention-forming way. A convention of trust, likewise, is a conventional regularity of having or forming a belief that p upon hearing someone else utter s. A grammar is correct if it is the best axiomatic theory among those which generate that correct language.

In more detail, a regularity R is a convention in a population P iff within P, the

---

[12] One author takes the fact that we can develop a successful endogenous version of Lewisian metasemantics as evidence for Lewisian metasemantics. However, it may be possible to develop endogenous versions of interpretationist and inferentialist metasemantics to rival the endogenous Lewisian account.
[13] Lewis 1983 pp163-188

following hold, with at most a few exceptions:

(1) Everyone in *P* conforms to *R*.

(2) Everyone in *P* believes that everyone in *P* conforms to *R*.

(3) This belief gives everyone in *P* a good reason to conform to *R* himself.

(4) There is a general preference in *P* for general conformity to *R* rather than slightly-less-than-general conformity to *R*

(5) There is an alternative possible regularity *R'* such that if it met (1) and (2), it would also meet (3) and (4)

(6) All of (1-5) are common knowledge.

To repeat, the specific conventional regularities that make an abstract language *L* used in population *P* are as follows, where *s* ranges over sentence types and *p* over propositions,

(Truthfulness) Members of *P* utter *s* only if they believe *p*, where *L*(*s*)=*p*.

(Trust) If a member of *P* hears another member of *P* utter *s*, she tends to come to believe *p*, where *L*(*s*)=*p*.

(Lewis) summarizes the metasemantic theory:

(Lewis) Given an exogenously fixed specification of population P_1 and typing of sentences, T_1, *L* is the language of P_1 for T_1 iff there are conventions of (Truthfulness) and (Trust) in *L* in P_1 for T_1.

This highlights that if (Lewis) is to provide a naturalistic metasemantic theory, it must be supplemented by exogenous theories of populations and word types.[14] However, we've seen reason to doubt that the available exogenous accounts of word types will serve. As before, an exogenous account with semantic elements is circular. As before, combining a metasemantic account with an exogenous account that cuts too finely, such as (phonographemic) causes havoc in the semantics. For example, in the "okapi" case, either

---

[14] It must also be supplemented by the contents of the beliefs of members of that population. Lewis' theory is then, 'headfirst': the intensional content of mental states is presupposed in the characterization of semantic content. The authors of this paper differ in their enthusiasm for granting this headfirst presupposition.

the convention covering those tokens is one that ties them to zebras or there are too few tokens of the 'word' for there to be any convention.

We will use the remainder of this section to show why overly coarse-grained accounts are problematic in combination with (Lewis). Because of the complexities and flexibility of Lewis' account, this will take some time. Fortunately, the work here will also be useful in defending our own account in section 6.

Overly coarse grained accounts of words obscure the differences between word tokens with intuitively different semantic values and thereby conflate what should be separate streams of input to the metasemantic theory. For example, tokens that are intuitively about financial institutions/trains are bundled in with those intuitively about waterway edges/drains. We focus on case of "bank" but similar reasoning could be applied to the "train"/"drain" case. Call the phonographemic string "There is a bank nearby" sentence type s_1. Which language and grammar can capture the conventional regularities for s_1 for English speakers? The problem is that sometimes a speaker produces s_1 when she believes there is a financial institution nearby; but other times she does so when she believes there is a river edge nearby. Sometimes hearing s_1 inclines a hearer to adopt the one belief and sometimes the other. Which grammars and which languages match those regularities while treating all instances of the phonographemic string as, counterintuitively, a single sentence type? A grammar that associates "bank" with only financial banks will not suffice, nor will one that associates it with only river banks. There are three ways that Lewis could address this problem: the disjunctive method, the sequence method and the indexical method. We'll show none of them renders the phonographemic account suitable as an exogenous plug-in for Lewis' metasemantics.

We start with the *disjunctive method*. Let the grammar G_d pair "bank" with a function from the union of the set of financial banks *and* the set of river banks to true. So, each sentence containing "bank" is paired with a belief that has disjunctive content. Suppose further that G_d includes the usual rules of composition. G_d pairs s_1 with a proposition of the form [there is either a financial institution or the edge of a waterway near to x] where x is a place provided by context. This might capture the regularities of English for s_1. However, such a theory will struggle to get the right regularities for phonographemic strings of the form "There are three banks". This is uttered when there are either three financial institutions or

three river banks; not when there are three things that are either financial institutions or river banks as the conventions of G_d require. In such cases, the disjunctive method delivers the wrong semantics. An analogous problem afflicted (Toy Evans).

Lewis' own approach to ambiguity is to pair each sentence s with a *sequence* of propositions *<p, q….>*, instead of pairing it with a single proposition. The truthfulness convention is a conventional regularity of uttering *s* only if one believes *p*, or believes *q*, or …. The trust convention is a conventional regularity of believing *p* or believing *q* or … if one hears another member of the population utter *s*. According to this proposal, "the bank is nearby" is paired with the sequence: *<the financial bank is nearby, the riverbank is nearby>*. Call this *the gruesome sequence regularity*. "There are three banks" is paired with the sequence *<there are three financial banks, there are three river banks.>* Lewis does not describe a grammar that generates such sequences. Perhaps he has in mind a grammar G_s that associates words with sequences of referents, with each sentence *s* then mapped to sequences of all those propositions that are determined compositionally from some selection of referents from the list associated by G_s with the constituent words of *s*. For example, "bank" will be associated with the list <financial bank, river bank>. Call this *the sequence method*. If G_s does pair "there are three banks" with the extensionally appropriate sequences, it avoids the problems of the disjunctive method.

The sequence method also fails. The problem is that the gruesome sequence regularity is not a convention. To be sure, there is a regularity of uttering "the bank is nearby" either when one believes that the financial bank is nearby or believes the riverbank is nearby, and a regularity of coming to believe one of those propositions when one hears the sentence uttered. But for the gruesome sequence regularity to be a convention, Lewis's condition (3) requires that our awareness of that regularity gives us reason to conform to it. It does not. Suppose you hear someone utter "the bank is nearby". Awareness of the gruesome sequence regularity gives you reason to believe that either the speaker believes a riverbank to be nearby, or believes a financial bank to be nearby. If you assume the speaker is reliable, this gives you reason to believe the disjunction: either a riverbank or financial bank is nearby. But believing the disjunction is not to conform to the gruesome sequence regularity of trust. To do that, one would have to believe one or other of the disjuncts.

Turn now to the *indexical method*. Here is one way of fitting indexicals into a Lewisian account.[15] Pair sentences with characters, functions from context to 'horizontal' content. While there's no horizontal proposition p such that every speaker will utter "I am standing" only if they believe p, all speakers satisfy the following: they will try to only utter "I am standing" (in c) when they believe whatever proposition which is the value of that sentence's character C at c; likewise, they will all try to respond to utterances of that sentence-type (in c) by forming the belief in the same proposition. An example of such a conventional regularity is: utter "I am standing" only if A is the speaker, and one believes that A is standing. Notice that a speaker *can* use awareness of this regularity to give them reason to believe an appropriate formulation of trust: for they combine their awareness of the general regularity *with the publically available information about which context they're in* (who the speaker is) and, if they take the speaker to be reliable, they'll have reason to believe that that individual is standing, and so to conform to the regularity.

One could try co-opting that treatment of indexicals to the case of "bank". On *the indexical method*, the convention would associate "bank" not with a sequence of propositions, but with a function from contexts to propositions. The putative regularity would be that a speaker would only utter "the bank is nearby" when *financial banks is the salient disambiguation of "bank"* and they believe that the financial bank is nearby or *riverbanks are the salient disambiguation of "bank"* and they believe the riverbank is nearby. Belief in these regularities combines with knowledge of context to provide reasons to conform to the analogous regularity of trust just as it does for standard indexicals. Call the relevant regularities *the gruesome indexical regularities*.

One problem for the indexical method is how to keep the base free of semantic facts in line with our reductive ambitions. The gruesome indexical regularities build in the "salient disambiguation of words" as part of context, where the account of context is an exogenous account plugged into the reductive metasemantic account. But what makes a disambiguation salient depends in large part on the linguistic context in which the word appears (e.g. whether mortgages or boating was mentioned most recently). "Salient disambiguation" is a

---

[15] A different interpretation of Lewis uses the (later) distinction between horizontal and diagonal content. (When A is the speaker, the horizontal content of "I am sitting" is that A is sitting; the diagonal content is that the speaker of the context is sitting). Diagonal contents of sentences are a candidate for the narrow content of beliefs ascribed using those sentences. This might well serve for true indexicals. But the analogue of diagonal content for *ambiguous* sentences would be highly recherché, and it's implausible that ordinary competent use of "bank" requires one to form beliefs with that content.

smokescreen for sneaking in illicit appeal to the semantic. Upon examination, this particular indexical account is no less circular than helping oneself to semantically-individuated words at the outset.[16,]

A problem for all three methods of dealing with overly coarse-grained words – disjunctive, sequence and indexical - is that, even if they identify regularities of truthfulness and trust for a population, these regularities are not conventions because they do not meet condition (3) – they do not reflect our reasons for continued conformity. We postpone explaining this problem until section 6.

The exogenous version of Lewis's theory—Lewis's own approach—is no more promising than (Toy Evans). We will now adjust the Lewisian metasemantics to provide a positive, endogenous metaphysics of word-types.[17]

<div align="center">3.</div>

Let a language be a triple of a population P, a typing relation T and a function L from sentence types to propositions. The population is a set of time-slices of people.[18] The typing relation is a set of sets $s_1...s_n$ where each s_i is a set of actual and possible token utterances. T should impose a typing on each of the utterances of members of P. L is a function from those sentence types $s_1..s_n$ and only those sentence types to propositions.

If languages are triples of this sort, how do the conventional regularities of truthfulness and trust fix which of those languages is the one in use for a particular utterance u? We replace (Lewis) with (Endogenous).

(Endogenous) Given an utterance u,

<P, T, L> is a language in use in utterance u iff P is a population and T a

---

[16] This is not to deny that one can specify other features of context non-semantically e.g. the speaker, time, place of the utterance.

[17] Of course, Lewis' approach may need other refinements. For example, Lewis considers objections to his approach relating to liars and non-literal utterances in his 1983 pp163-188. But in these cases, any responses open to Lewis are also open to our endogenous account. The aim of our paper is to only to explore an example of theory of words endogenous to some metasemantics and show it fares better than the exogenous version of that metasemantics.

[18] Populations must be sets of time-slices of people in order to allow division of the utterances of people who are bilingual.

typing relation relative to which there are conventions of (Truthfulness) and (Trust) in L, and the speaker/hearer of u is a member of the population P; and u is a member of some equivalence class of the typing relation T.

Instead of determining L *after* fixing a particular population and typing relation, (Endogenous) treats the population and typing relation as variables whose values are fixed however is necessary to produce conventions of (Truthfulness) and (Trust).[19]

If P is to count as language-using population, the members of P have to be able to think about population P, and think about the sentence types T induces, in order to have the belief that the relevant regularities prevail among P. This is not a problem for (Endogenous), as the way that speakers think about P and about S is fairly unconstrained. For example, suppose that each member has an unstructured name-like concept ("Us") and each believes that every one of Us follows the relevant regularity. Suppose also they have an unstructured concept "sentence-of" whose referent is a function from utterances to types, which features in the specification of the relevant regularities. What the account requires, in order for $<P,T,L>$ to satisfy (Endogenous), is that these unstructured concepts "Us" and "sentence-of" refer to the right things (P, and the function induced by T, respectively). Why be pessimistic about this issue? One *might* think it problematic if one assumed that members of P think about P and T descriptively, via some implicit exogenous theory of populations or sentence-types. But why restrict speakers to descriptive concepts of P and T?

4.

Notice that (Endogenous) makes space for languages where words are typed so as to have the extension specified by the phonographemic or Kaplanian accounts, just as long as those types do in fact feature in conventions of the right sort. The considerations earlier are reasons to think they won't feature in suitable conventions for languages such as English. But nothing in the *form* of the endogenous account rules them out.

The endogenous account is highly flexible, making room for words to be typed so as

---

[19] While population and word-typing are fixed endogenously by our new metasemantic account, we still presuppose that there are exogenous naturalistic accounts available for other entities, in particular, the sounds and marks that comprise the utterance *tokens*.

to avoid the overly coarse or overly fine graining we saw in extant exogenous accounts. (Endogenous) makes room for word types to have the extensions which (Plus Semantic) and the indexical method were reaching for (again, provided those groupings of tokens feature in some conventions of the right sort). Nothing prevents us from giving an informative specification of the extension of word-types that appeals to semantic facts, so long as this is not construed as part of a reductive metaphysics.

In particular, (Endogenous) makes room for a satisfactory treatment of cases that phonographemic typed too finely. On (Lewis), the grapheme "realise" needs a different convention from the grapheme "realize" since words are typed phonographemically. By contrast, (Endogenous) permits a typing that groups these together as tokens of one word. (Endogenous) also addresses the problem of word-types with too few instances to support conventional regularities. The single pronunciation of "okapi" in a North Welsh accent can be lumped together with pronunciations in other accents.

What about cases that the phonographemic account typed too coarsely, such as the "bank" case? (Endogenous) allows the language to feature a typing relation T that types some utterances such as B-A-N-K as of the word type "bank$_1$" and others of the type "bank$_2$". With two words, the language can have a grammar featuring a non-disjunctive regularity, pairing "bank$_1$" with the function from riverbanks to True and a separate non-disjunctive regularity pairing "bank$_2$" with the function from financial banks to True. In a case where there are at most two riverbanks and one high street bank, "there are three banks" will come out false as desired, avoiding the problem that faced the disjunctive method. (Endogenous) also avoids the problem that faced the sequence method. There are regularities of truthfulness and trust governing respectively the types "there is a bank$_1$ nearby" and "there is a bank$_2$ nearby", speakers utter the first only when they believe there's a riverbank nearby, and the second when they believe there's a financial bank nearby, and form the appropriate beliefs when they hear the respective utterances. A speaker's reason to obey (Truthfulness) for the sentence featuring bank$_2$ is her knowledge that her interlocutors conform to (Trust) for that sentence. So unlike the sequence method, the regularities in (Endogenous) meet condition (3) for conventions. Finally, although the typing reflects semantic features of the token sentences, this does not result in circularity. The indexical method was embroiled in circularity only because it was part of an exogenous theory. Since we do not attempt to characterise T prior to the semantics, there is no circularity.

(Endogenous) also makes *populations* endogenous to the metasemantic theory. This answers an objection White (ms) raises against Lewis' metasemantics. White gives reason to doubt that there is a principled non-semantic exogenous account of a population. The speakers of a language are not exactly those within certain geographical boundaries in certain time periods or those of certain nationalities. Nor can we identify the population as those who speak a certain language if the population is to be given prior to the semantic theory. Our endogenous approach gives a principled line on what will count as a language-using population: a language speaker population is any group for which there are conventional regularities of truthfulness and trust. So our approach rescues Lewis from White's objection.

5.

Although we have made room for new typings of utterance tokens, this will be no good if it comes at the cost of removing too many constraints, leaving us with an account that lets in crazy typings along with the desirable ones, thereby letting in crazy semantics along with the desired semantics. In this section, we suggest *prima facie* ways to gerrymander such crazy languages. Fortunately, in section 6, we can rule out that these gerrymanders are genuine languages - although they may characterise genuine *regularities* of truthfulness and trust, those regularities are not *conventions*.

We will consider three ways our opponents might try to gerrymander spurious languages from intuitively correct semantic theories – by restriction, by merging, by tailoring. Begin with restriction. If a regularity prevails among the whole population, it also prevails among subpopulations. One can gerrymander spurious languages (it seems) by taking the intuitively correct semantic theory and restricting the population. The restricted subpopulation could be arbitrary, or it could be based on some recognizable feature. For an example of the latter, take English but restrict the population to the brown-eyed subset of the original population (counting those with other colours of eyes as a separate linguistic population). Relatedly, there are ways to subdivide the *typing* relation while preserving regularities of truth and trust. For example, take English but double the number of types of word, counting whispered and non-whispered tokens always as of different words.

A second method to produce gerrymandered types and languages is merging. One trick is to merge what are intuitively distinct but synonymous word types into a single gerrymandered word type. For example, type tokens of the graphemes "cell" and "mobile" together as if they were different graphemic realisations of the same word. Another gerrymander merges tokens of the graphemes "purchase" and "buy" together. Merged gerrymanders keep the structure of regularities in place, simply collapsing two regularities from intuitive language(s) into a single regularity in the merged gerrymander. For that reason, merged gerrymanders would produce the same meaning assignments as the intuitively correct semantic theory that they are based on.

Another trick merges tokens which are intuitively of distinct non-synonymous word types. An example groups all tokenings of "there are oranges in the fruit bowl" together with one person's tokenings of "there are fish in the sea", counting these as instances of a single type. A more natural version would count tokenings of the same phonographemic type as of the same word-type across geographical or historical semantic drift. For example, "pants" picks out an external leg-covering among speakers from one geographic region, an undergarment among those from another.

How should non-synonymous mergers specify the candidate regularity of truthfulness/trust to cover disparate usage of the two non-synonyms? To get the right semantics, one would have to treat the merged 'word type' as ambiguous. We discussed three methods of dealing with ambiguity in section 3 – the disjunctive, sequence or indexical method. (In essence, treating homophones like "bank" as a single type is also a merge relative to the two non-synonymous but phonographemically identical types.). We showed in section 3 that the disjunctive and sequence methods are unsuitable. However, we could treat merged words as indexical provided the relevant aspects of context are not semantic as we argued they were for homophones like "bank". For the *orange/fish merge* the relevant feature of context is whether the token was uttered by a particular person and this is not problematically semantic. For "pants", that contextual feature might be the accent of the speaker.

Restriction and merging (under the indexical treatment) in general do not generate unwanted assignments of truth-conditions to any token utterances. By contrast, our final kind of spurious typings - tailored typings - don't only produce spurious types; they also threaten to overgenerate meaning assignments to tokens. Tailoring involves tweaking the boundaries of standard types, so as to alter the semantic assignments to particular utterances.

Arbitary tweaking is not guaranteed to produce regularities. But here's a prescription for gerrymandering genuine regularities: find utterances which are exceptions to the regularities of what is intuitively the correct semantics, then adjust the population or sentence typing so that they are no longer exceptions to the rule that covers them.

Here is a first example of tailoring - *the red tailor*. Consider a biased selection P* of the population who are apt to call more orangey things "red" than is the norm. Suppose P* are scattered among the rest of the intuitive population of English speakers. Now consider a gerrymandered typing of words, "red$_1$" and "red$_2$", the first tokened exclusively by members of P* when interacting with other members of P*, the other tokened on the remainder of the occasions. By construction, there will be regularities of truthfulness and trust among P* linking "that is red$_1$" to a belief that the item in question is red*, where red* is a colour that includes orangey non-red shades. Insofar as we concentrate on the whole population and the standard typing of "red", the "red" tokens uttered by members of P* will mean *red*; but insofar as we concentrate on the gerrymander P*/ "red$_1$" those very tokens will also mean *that is red** (in a distinct language).

For a second example, consider utterances where the speaker intended to produce the phoneme /treɪn/ but the phonetic intention failed and she produced a different phoneme, /dreɪn/. Perhaps many such utterances are misheard as /treɪn/ by the audience, a common mistake, given expectations set up by the conversational context. Call these *the fluffed tokens*. Intuitively the fluffed tokens are instances of the word "drain", not "train". Accordingly, the standard semantic theory treats sentences involving the fluffed tokens as exceptions to the truth and trust regularities linking "drain" and beliefs about drains. (Lewis) and (Endogenous) permit this - regularities need not be perfect. But our opponents can retype the fluffed tokens (which sound /dreɪn/) with tokens that sound /treɪn/. Call this word-type "train+". There are regularities linking "train+" to beliefs about trains, which both the tokens sounding /treɪn/ and the fluffed tokens confirm.

Malaprops too are susceptible to tailoring. Suppose Mrs Malaprop says, "Aviators are dangerous reptiles found in the marshes in Florida." On the intuitive semantic interpretation, Mrs Malaprop says something false about people who fly planes, even if her audience can work out what truth she intended to convey. Our opponents tailor a gerrymander – *the malaprop tailor* - out of Malaprop's exception to the intuitive regularities. Retype the first

token word in Malaprop's utterance with tokens spelt and sounded like the intuitive word-type "alligator" to produce a word type "alligator+". One can now subsume her utterance under a truthfulness regularity linking alligators and "alligator+". Perhaps Lydia hears Mrs Malaprop and takes her to believe that there are thousands of alligators. If so, one can also subsume Mrs Malaprop's utterance under a trust regularity.

<div align="center">6.</div>

We have presented seven ways of gerrymandering word-types. There were two forms of restriction: either arbitrary (subtypes that do not correspond to features to which speaker-hearers are sensitive) or recognizable (whispered vs. non-whispered, uttered by someone with brown vs. blue eyes); two forms of merging (of synonyms, e.g. "cell"/"mobile", "buy"/"purchase", or of non-synonyms under the indexical treatment); and three ways of tailoring the type (the red-tailor, favouring orange-biased usage, the fluff-tailor, reclassifying drain-utterances as train-utterances, and the malaprop-tailor, treating Malaprop's aviator-utterance as an alligator-utterance). These all articulate genuine regularities of truthfulness and trust connecting our utterances and attitudes.

Fortunately, Lewis' account of convention already contains within it countervailing pressures to rule out these gerrymanders. To be a convention, a regularity needs to do more than meet clause (1) of Lewis's definition. Condition (3) requires that the populations' beliefs that the regularity obtains must give them good reasons for conforming to it in future. We will argue that this means that types must be (i) identifiable; and (ii) must be psychologically present.

	Restriction		*Merge*		*Tailor*		
	arb.	*recog.*	*syn.*	*non-syn.*	*red*	*fluff*	*malaprop*
Identifiable	N				N		
Presence	N	N	N	N	N	N	N

The table summarizes what these constraints rule out. Although psychological presence would suffice alone to rule out all gerrymanders from counting as actual languages, identifiability tells us something about what is a possible human language. For this reason it can explain why the phonographemic account of words seemed appealing.

Identifiability. To feature in a conventional regularity, sentence types and populations must be identifiable by speakers, in the sense that the speaker must have a generic capacity to tell that utterance *u* is of type *s*, or person *a* is in population *P*. This falls out of (3)'s requirement that speakers employ beliefs about the type in reasoning. (3) requires that the speaker-hearers' awareness of the existence of the regularities of truthfulness and trust prevailing in their population give them reason (either epistemic or practical) to conform to those regularities. Here, for illustration, is the schematic form of the epistemic reason to conform to a suitable instance of the trust regularity:

a. There's a regularity among *P* to utter type *s* only if they believe *p*.
b. The speaker has uttered something of type *s*.
c. The speaker is a member of *P*.

So,

d. The speaker believes *p*.

Combined with standing presumptions of cooperation and expertise, this supports a conclusion of the form:

e. *p*.

(Similar reasoning applies on the practical side.) For this sort of reasoning to be available to a hearer, she needs, in steps b and c, to identify which type a token she encounters falls under, and which population a person stage she encounters belongs to. If she can never do this, her belief in the truthfulness regularity never provides her with reason to conform to the trust regularity and (3) fails utterly. To be able to identify types and populations, she need not do so successfully in all cases. For example, Zara misidentified Xena's utterance of "train" but she still has the general ability to identify utterances of "train". However, given that in order for a regularity to be convention, (3) must be met *with only a few exceptions,* she must be able to identify types and populations *with only a few exceptions*.

The identifiability point is general to all conventions. Take the *convention* of driving on the left, when in Australia. There is therefore a regularity of driving on the left when in region *R* of Australia (and one of driving on the left when in the Australian complement of *R*). But if I can't tell when I'm in region *R*, and when not, these subregularities can't feature in my practical reasoning and if so, these regularities aren't *conventions*.

The identifiability constraint rules out some of the gerrymanders from section 5. It immediately implies that *arbitrary* subpopulations (P_1/P_2) or *arbitrary* finer-grained typings of words (s_1/s_2) won't feature in conventions, since speakers will not in general be able to identify when an utterance falls under s_1 rather than s_2, or an individual is in P_1 rather than P_2. The constraint also eliminates some examples of tailoring: speakers can't tell who is biased towards orange in their "red" grapheme/phoneme utterances if P* is unidentifiable. As the table records, many gerrymanders are not ruled out by this constraint. Recognisable restrictions, the malaprop and the fluff tailor and the various forms of merging, all pass the constraint.

The identifiability constraint does admit the intuitive sentence-types and populations, as desired. The population of English speakers make distinctive noises and inscriptions that enable others in the group (with only occasional exceptions) to identify them as such. Tokens of the intuitive word type "cat" are also identifiable as such, with the primary clue being their phonographemic features. Background knowledge plays an important mediating role in this identification: phonological features are combined with experience of the idiosyncratic variations of pronunciation across different accents or speech impediments to produce our ability to identify types. Other resources available to us in the identification task include non-semantic conventions[20] such as the explicit letter transformation conventions of braille and transliteration conventions, both of which apply regardless of semantics to any sentence type. This explains the initial appeal of an exogenous phonographemic account. Phonographemic features are indeed important but they feature in the epistemology of type recognition, rather than the metaphysics of those types. In the epistemology, they need only be defeasible evidence of word type. This comfortably accommodates the holism and idiosyncrasy that blocked a metaphysical phonographemic account.

Psychological presence. We take it that to be a convention, a regularity must be the content of the *psychological* reasons that members of a relevant population possess for acts of conformity to R.[21] Given this, a genuine regularity R* can fail to meet (3) if it does not feature as the content of speakers' reasons.

---

[20] See Cappelen 1999 for discussion of such conventions.

[21] An alternative reading of (3), which we reject, is that the regularity should be a consideration that objectively counts in favour of future conformity to it, irrespective of whether it is psychologically present in the population. The motivation for the objective reading is to avoid over-intellectualizing speakers by attributing to them explicit beliefs about the regularities. However, the psychological view need only attribute implicit beliefs.

Again, there are illuminating precedents in driving conventions. First consider merging: From the base conventions, such as driving-left-in-Australia, we can construct merged regularities such as driving-left-when-in-Australia-and-driving-right-when-in-the-US. These merged regularities do not feature in agents' psychological reasons for instances of left-driving in Australia. As evidence, notice which information agents take as relevant when deciding which side of the road to drive on. The reason for and decision to drive on the left in Australia are resilient under various hypotheses about what happens elsewhere. I don't recheck my decision to drive-left-in-Australia if I discover that the US has altered its driving laws. This shows that the Australia specific regularity is operative in my reasons, not a merged regularity for both Australia and the US.

Exactly the same goes for mergers of types for conventions of truthfulness and trust. We psychologically encode information about truthfulness and trust for tokens of "buy" and "purchase" separately. As evidence, notice that were I to learn that the meaning of "purchase" had shifted so as to be associated with the belief an item had been stolen, I would not regard my practice with "buy" as undermined. *A fortiori,* arbitrary mergers of non-synonymous words, such as "orange" and "fish" do not figure among my reasons for forming beliefs or uttering sentences. Equally, in a conversation between Americans, it doesn't matter whether or not they know the British usage of "pants". That shows that the US specific regularity rather than the merge of US and UK tokens of P-A-N-T-S is psychologically present. We now have an additional reason to reject Lewis' treatment of ambiguity from section 2. It is the two separate bank$_1$ and bank$_2$ regularities that feature in our psychological reasons, not the disjunctive, indexical or sequence regularities featuring the phonographemic type B-A-N-K. Evidence is that the former but not the latter regularities are resilient. If I get evidence that the regularity linking B-A-N-K to river banks is breaking down, it affects only my watery uses of B-A-N-K, not my financial ones. In sum, *presence* knocks out the identifiable but gerrymandered mergers.

Application of the presence constraint also rules out recognizable restrictions. Again, consider the driving convention analogue. From a base convention of driving-left-when-in-Australia, we get restricted regularities of driving-left-when-in-a-big-car-in-Australia and driving-left-when-in-a-small-car-in-Australia. Here, the relevant psychological fact is that we *are* inclined to recheck our practical reasoning when in a big car, conditional on learning that behaviour with small cars is not as we thought it was. If our psychological reasons for drive-left (in circumstances where we happen to be in a big car) were the big-car regularity, this

should not happen.

The whispered/non-whispered restricted regularities are the analogues of this in the case of truthfulness and trust. If I learn that whispered tokens of "blue" are applied to slightly greener things than I had thought, I'm inclined to assume that this generalizes to non-whispered tokens. We're disposed to take it that whispered and non-whispered tokens pattern alike---being more confident in this than we are in the exact patterns connected to either. So our actual psychological reasons for the linguistic acts (and beliefs formed in communicative situations) are not sensitive to the whisper/non-whisper distinction.[22]

It remains to consider the fluff and malaprop tailors. The underlying challenge here is to pinpoint why paradigm conventions are based on simple but exception-prone regularities, when there are complex regularities with fewer exceptions that we are in a position to work out.

The Australian regularity driving on the left has exceptions, predictable from our general knowledge of the world and each other: we violate the regularity when swerving to avoid a Kangaroo, when moving into a parking space on the right, etc. One could tailor a "regularity" to which there is more perfect conformity by including the exceptions into the specification: "drive on the left except when one swerves to avoid a kangaroo,…". The latter gerrymander doesn't seem to be a true convention. Why not?

We can once more appeal to presence. The simple rule remains a reason for us to drive left, conditional on varying assumptions about the character of exceptions. We might learn that drivers do not react fast enough to swerve to avoid a kangaroo. That is a useful piece of information, but not something that prompts me to recheck my ordinary practical reasoning about how to drive in the absence of kangaroos. When kangaroos are around, the complex regularity does no better, for it is *not* a psychological reason for swerving. When I do swerve the reason to do so is that if one doesn't one will have a collision, not a desire to conform to some rule.

---

[22] This is not to deny that there can be nested conventions, cases where the inner conventions are restrictions of an outer, more general convention, perhaps for a larger population. For example, there seem to be specific conventions for speakers with a Leeds accent nested within more general ones for speakers with a British accent, nested inside more general conventions still for speakers with a wider range of accents. These nested conventions might feature different types – consider the case of "colour"/"color".

In the case of truthfulness and trust, malaprops and fluffs pattern similarly. Regularities featuring the type "alligator+" have one fewer exception than the 'standard' regularities featuring the intuitive word types. However, that gerrymandered typing is not anybody's reason for uttering what they do. It's not Mrs Malaprop's, since she is not aware of her idiosyncracy. It's not our reason when interpreting each other, outside malaprop contexts. We don't recheck our way of interpreting each other conditional on propositions about that utterance of Mrs Malaprop's. You might think that knowledge of the malaprop regularities would be an interpreter's reason for attributing to Mrs Malaprop an alligator-belief upon hearing her utter, "Aviators are dangerous reptiles found in the marshes in Florida." But anyone who tried to appeal to that rule would make a mistake about Mrs Malaprop's psychology. They would be representing her utterance as the rational outcome of *true* beliefs about regularities in usage of a (tailored) word-type, when part of the basic data about this case is that she's either subject to performance-error (so that act-type isn't rationalized by her beliefs) or has false linguistic beliefs that lead to her utterance. Similar remarks cover the case of fluffs. So, the psychological presence constraint disqualifies all the remaining gerrymanders from being languages. The threat of overgeneration is allayed.

There's a difference between regularities that are not conventions due to violations of identifiability, and regularities that are not conventions due to violations of presence. The former regularities *could not* become conventions, unless our powers of recognition were enhanced. In the case of presence-violators, it is a contingent fact of psychology that we reason one way rather than another. We *could* base our driving decisions on international disjunctive regularities rather than national ones. We *could* treat "buy"/"purchase" as a semantic unit, a variant spelling like "recognize"/"recognise". One shouldn't expect the abstract metaphysics of words to be the place where an illuminating explanation of why our psychology of language codes things one way rather than another---historical linguistics is a much better bet for such insight. But our point has been simply to show that the endogenous story set out in section 3 was not immediately refuted by crazily overgenerating word types, and this has been achieved.[23]

<center>7.</center>

We have set out and defended a new account of sentence-types. Rather than receiving

---

[23] One author would also use condition (4) of Lewis's characterization of convention to demonstrate the non-conventionality of some of the gerrymandered word-type regularities.

some individuative story prior to and independent of our metasemantics, on our theory the same factors that fix meaning, fix word-typing. In so-doing, we avoid the problems that beset previous accounts of the metaphysics of words, including the widely but unreflectively endorsed phonographemic view. Our story explains why the phonographemic view seemed so natural: phonographemic factors play a crucial epistemological role in the *identifiability* of our sentence-types.

The most important worry for our view is that it might overgenerate spurious word-types. We have explained in detail why, although there are many regularities of truthfulness and trust involving spurious typings, there are very few *conventional* regularities.

We have given an account of *sentence types*; it is a further question what individuates the *lexical items*, the basic morphemes out of which sentences are built. A natural extension of our account addresses this question. Recall that Lewis held that once we had a grip on which language (function from sentence-types to propositions) was in use in a population, we would then use this to get a grip on which grammar (syntactical parsing and semantic analysis) was in use therein. The grammar in use was the simplest, strongest theory that generated the sentence-proposition pairings of a language-in-use. Included in the grammar, in our view, will be an assignment of lexical-types to parts of utterances, as well as an assignment of syntactical-tree types to whole utterances; a constraint will be that the leafs of the tree be lexical types that (if non-null) are instantiated by parts of utterances. The requirement for simplicity and strength will ensure that a sensible typing of sentence-types will require a sensible typing of constituent word-types, if the grammar is to be used.

Although our focus has been on the metaphysics of word-types, we have *en passant* given an account of language-using populations. Thus we are able to offer an answer to White's (*ms*) challenge: that the Lewisian metasemantics is circular because the relevant populations cannot be non-circularly specified. White presupposes that the account of populations must be exogenous.

The endogenous theory of word-types allows the naturalistic theory of meaning to proceed without the worry that the metasemantic base contains tacit circularities. But the flipside is that there is no guarantee that endogenous theories are available for arbitrary metasemantics. Those wishing to deploy a sophisticated variant of the Evans-style dominant-causal-role metasemantics with which we began still owe us all a (presumably exogenous) account of the metaphysics of word-types. Worse for them, the convention-based competitor

has discharged their debts in this respect---it can no longer be dismissed as *everyone's* problem.

<u>Bibliography</u>

Cappelen, H. 1999 "Intentions in Words", *Noûs*, 33:1, pp. 92–102

Cappelen, H. and Dever, J. 2001 "Believing in Words", *Synthese*, Vol. 127, No. 3 (Jun., 2001), pp. 279-301

Davidson, D. 1973 "Radical Interpretation", *Dialectica*, 27: 314–28

Davidson, D. 1979 "Quotation," *Inquiries into Truth and Interpretation* (New York: Oxford, 1979)

Evans, G. 1973 "The Causal Theory of Names,'" *Proceedings of the Aristotelian Society, Supplementary Volume* 47: 187–208

Haack, S. 1978 *Philosophy of Logic* (New York: Cambridge, 1978)

Hawthorne and LePore 2011 "On Words", *Journal of Philosophy* cviii, no. 9 pp. 447-485

Horwich, P. 1998 *Meaning* (OUP, 1998)

Kaplan, D. 1990 "Words", *Proceedings of the Aristotelian Society (Supplementary Volume),* 64: 93–119.

Larson, . and Segal, G. 1995 *Knowledge of Meaning: An Introduction to Semantic Theory* (Cambridge: MIT Press, 1995)

Lewis, D. 1983 "Language and languages" in *Philosophical Papers, Volume I,* (Oxford: Oxford University Press 1983) pp163-188

Millikan, R.G. 1984 *Language, Thought and Other Biological Categories* (Cambridge: MIT Press. 1984) pp72-6

Richard, M. 1990 *Propositional Attitudes: An Essay On Thoughts and How we ascribe them*, (Cambridge, Mass: Cambridge University Press 1990) pp. 183-90.

Stebbing, L. S. 1935 "Sounds, Shapes, and Words", *Proceedings of the Aristotelian Society (Supplementary Volume* 14 (1):1-21.

White, R. (ms). "David Lewis, meaning and convention."