Near-final version of paper that appeared in Nous Online in April 2016

Egocentric Content

Hartry Field
New York University and University of Birmingham

Abstract: The paper distinguishes two approaches to understanding the representational content of sentences and intentional states, and its role in describing people, predicting and explaining their behavior, and so forth. It sets forth the case for one of these approaches, the "egocentric" one, initially on the basis of its ability to explain the near-indefeasibility of ascriptions of content to our own terms ("'dogs' as I use it means *dogs*"), but more generally on the basis of its providing an attractive overall picture of the descriptive and explanatory role of representational content. In doing this, the paper relates the egocentric view to an "immanent" or "deflationary" view of reference and truth conditions, and also to the view of reference-talk and truth-talk as anaphoric devices. It discusses the indeterminacy of content ascriptions to those in communities with radically different theories, a phenomenon that is unsurprising on the egocentric approach, and connects this to the thesis of the normativity of meaning. (It does all this in rather broad brush: many strands of the egocentric account will be familiar, and are the subject of familiar controversies; the point of the paper is less to address these controversies individually than to tie the strands together into what is hoped to be an appealing package.)

In finding out about the world, one thing we make use of is the beliefs of others. I see someone pick up her umbrella before she leaves the apartment, and conclude that she has formed a belief that it is fairly likely to rain during the day; this may increase my own confidence that it will rain. (Especially if I know that she usually consults weather sites on her computer before going out, sites that are reasonably reliable.) Or I hear her curse while cooking in another room, and conclude that a minor disaster regarding dinner has probably occurred—an inference that goes via the assumption that she probably wouldn't have cursed had she not believed that a disaster had occurred (and would have done so much more loudly had it been major), and that on such matters her beliefs are usually correct.

It is clear that in such examples, it is essential to think of beliefs as having representational contents, in some sense of that term. From her picking up the umbrella I concluded that she probably believed *that it was fairly likely to rain*, and from that, I was led to raise my own degree of belief that it would rain. Perhaps in these very simple examples I could eliminate the middle man, and infer rain directly from her picking up the umbrella? Well, *maybe* so in these simple examples, but in more complex examples the task seems hopeless; moreover, if the suggestion could be carried out it would involve not merely eliminating content from the explanation, but eliminating belief from the explanation as well. It is hard to see how we could make much use of a notion of belief that didn't describe beliefs via representational content. In some sense of 'representational content'.

Of course, representational content is used not only in describing beliefs, but also in describing hopes, fears, suppositions, idle thoughts, and many more such things; and these too would figure into any serious account of how we learn about the world using other people. In

what follows, talk of belief will simply serve as an important example. The main question will be how to understand representational content. (I'll generally drop the term 'representational'.)

One obvious fact is that when we ascribe content in a language, we use sentences of that language (or sub-sentential expressions of that language, for certain mental states other than belief, like love and hatred). For a foreigner (or an animal) to believe something, they may need to stand in a certain relation to a representation in their own language or internal system of representation; even so, when we monolingual English speakers ascribe beliefs to them, we do so via English sentences, rather than via sentences of the language (if any) of the believer. That's what's involved in the *public* ascription of belief. What about if we merely *believe* (but don't say) that a non-English speaker believes that someone is at the door? In that case, maybe no *sentence of English* plays any role in our belief state; but it would seem that we employ some sort of *internal representation* (or *internal way of thinking*) of someone being at the door. So it seems that to privately or publicly attribute a specific belief to another (whether a person, or dog, or whatever) requires some way of ourselves representing the content of the attributed belief. It isn't a big leap to the conclusion that we use our own (public or internal) representations as standards in attributing beliefs (and hopes, fears, idle wishes, etc.) to others.

This isn't to say that we need to think that another's belief state is well-captured by any representation that we might give. Maybe the cat in some vague sense believes that the mouse will turn left, but 'the mouse will turn left' imposes conceptual distinctions that the cat doesn't have, and I doubt that there's a better way to ascribe the belief. In attributing that belief to the cat I engage in "dramatic idiom" (Quine 1960, p. 219): I use my sentence (or my internal representation, in the case of a private attribution) to formulate the content of the cat's state, despite its evident inadequacy. (And this dramatic idiom seems essential to ordinary predictions

and explanations of the cat's behavior.) For people too, it may be that "some of their concepts are very different from ours", in the sense that any attempt to formulate a belief of theirs "involving those concepts" is inadequate. Still, the only handle we may have on their contents or concepts is by contents or concepts of our own which to some degree or other approximate them. (Perhaps we can sometimes describe their concept by stating its relations (e.g. inferential relations) to others of their concepts. But for this to help we need to understand those other concepts, and that will be by approximating them with concepts of our own.) Approximate similarity of content is a notion that is bound to play an important role in an account of how we attribute belief and how we make use of belief attributions in finding out about the world.

At this point there are two apparently different ways we might go. One is to theorize about contents directly, independently of any sentences or internal representations that represent them; and also theorize about the representation relation that sentences or internal representations bear to these independently conceived contents. The other is more egocentric: start from our own representations, which as I've noted are used as standards in attributing contents to others; and view all talk of content as some kind of projection from this. That's vague, but I hope to clarify it as I go on. What we end up with on this second route needn't be conceived as incompatible with what we'd end up with on some version of the first route, but it at least will suggest a different take on the first route than what otherwise might have seemed natural.[1]

---

[1] The approach that follows borrows heavily from Chapters 4 and 5 of Field 2001, including especially the Postscript to Chapter 4. But the presentation in those chapters obscured some central points, especially regarding truth, where the decision to take a use-independent notion of truth as basic and define a use-dependent one from it (rather than, as below, taking the use-

1. Ordinary attributions of content. We ascribe content to intentional states (*intentional content*). We also ascribe content to utterances (*linguistic content*: what the sentence uttered means on a given occasion of use). What are these contents that we ascribe? I'd like to leave this question open for now (and leave open too whether linguistic contents and intentional contents are the same), but ask instead how we standardly represent contents. In ordinary life, we do this by expressions of our own language. The content of a typical German utterance of 'Hunde' is *dogs.* The content of a typical utterance of 'Hunde bellen' is that *dogs bark*.

In the case of declarative sentential content ('Hunde bellen'), the italics were simply playing the role of emphasis. Even without the italics, the presence of 'that' is enough to indicate the special role that the sentence 'dogs bark' plays in the content-attribution: its role as a content-indicator. Unfortunately, we have no analog of 'that' for content-attributions for sub-sentential expressions (e.g. 'Hunde'); there the use of italics serves to avoid the appearance of ungrammaticality. It's convenient to introduce an artificial notation that subsumes both sentential and sub-sentential cases; one such device is "linguistic content marks", as in:

(1) (A typical use of) 'Hunde' has the linguistic content $^{<<}$Dogs$^{>>}$

and

(2) (A typical use of) 'Hunde bellen' has the linguistic content $^{<<}$Dogs bark$^{>>}$.

These are intended to be simply a convenient rewriting of the claims above, in more uniform notation. (In the sentential case, you can call linguistic contents 'propositions' if you like; but

---

independent one to be essentially a special case of an egocentric use-dependent one) has seemed to many readers a point of substance rather than merely a matter of presentation. I believe that what follows more successfully highlights the main ideas of the egocentric approach.

there's a danger if we do so of incorporating controversial philosophical assumptions without argument, and we need a word for the sub-sentential linguistic contents too.)

I'm tempted to rewrite these as

(1') 'Hunde' (as usually used) means $^{<<}$Dogs$^{>>}$

and

(2') 'Hunde bellen' (as usually used) means $^{<<}$Dogs bark$^{>>}$,

and to speak of the double-brackets as meaning marks. However, this could cause confusion, in light of an ambiguity in the notion of meaning that comes out when indexicals are at issue. (I'll mostly ignore indexicals in this paper,[2] but it's worth commenting on this nonetheless.) In one sense of 'meaning', the meaning of 'I hate that' is the same from one occasion to another; but in another respectable sense ("the meaning the sentence has on a given occasion", *aka* "the proposition expressed on that occasion", *aka* "the linguistic content of the utterance") the meanings vary. Meanings in the first sense are, roughly, functions from contexts to meanings in the second sense. I think the 'means that' locution generally goes with the second sense: we don't normally say "The sentence type 'Ich hasse dass' means that I hate that", but rather "When he uttered 'Ich hasse dass', he meant that he himself hated the broccoli in front of him". So provided that we understand meaning in this second sense, I'm fine with (1') and (2') and their analogs for indexicals, and with calling linguistic content marks *meaning marks*; and I will occasionally lapse into this usage for convenience. But because of the potential for confusion with the other sense of meaning, it's safer to talk of linguistic content, and I will mostly do so.

---

[2] Well, 'Hunde bellen' presumably has an indexical element, the present tense, but I'll be ignoring that.

Intentional contents too are ordinarily specified by expressions of our language. We speak of someone dreaming, imagining, believing or hoping that giraffes bark, which we might formally represent in analogy with the above as dreaming, imagining, etc., $^{<}$giraffes bark$^{>}$. (In using single brackets for intentional contents and double brackets for linguistic, I don't mean to assume that these are different; I merely want to avoid pre-judging that they are the same.) I think it harmless to say that dreaming $^{<}$giraffes bark$^{>}$ is being in a state with the content $^{<}$giraffes bark$^{>}$. It is not unnatural to suppose that such a state has a "feature" (perhaps a "component", though that term seems more loaded) with the content $^{<}$giraffes$^{>}$ and another with the content $^{<}$bark$^{>}$; I don't want to take a stand on this now, but if so it motivates the use of intentional content marks that apply to sub-sentential and sentential alike.

As noted above, these representations of linguistic and intentional content may be somewhat inadequate for describing the beliefs of those who think in sufficiently different ways than we do and the speech of those in different cultures; but we're putting that aside for now.

A case where they are adequate is in content claims about our own language—or more cautiously, our own idiolect. It seems completely correct to say:

'Dogs bark' (as I use it) has the content $^{<<}$dogs bark$^{>>}$ (i.e., has the content that dogs bark)

and that

'Dogs' (as I use it) has the content $^{<<}$dogs$^{>>}$.

(If there is any ambiguity or penumbral shiftiness in 'dogs bark' or in 'dogs', imagine the quotation name restricted to the class of tokens that function in the same way as the tokens used

inside the meaning marks.[3])

An important question: *How do I know* that 'dogs bark' (as I use it) has the content <<dogs bark>> (i.e., the content that dogs bark), or that 'dogs' (as I use it) has the content <<dogs>>?[4]

The question may seem a silly one, but it isn't immediately obvious why it should be silly. After all, the quotation mark names refer to linguistic expressions, and the content mark names apparently refer to contents; we're asking how we know that a given expression of our language has a given content. That seems like substantive knowledge. It isn't silly to ask how we know that the German term 'Hunde' has the content <<dogs>>, so why should it be silly to ask how we know that the English term 'dogs' has the content <<dogs>>?

One "answer", if one can call it that, is: we know such things by means of a faculty that allows us to know the contents of our own words. Let's hope we can do better than that.

Here's a first serious attempt to answer the question:

It is part of learning to use the word 'content' that we come to accept the schema

---

[3] I take it that this is to be spelled out in terms of the processing that produced those tokens; very crudely, whether two of my tokens of 'bank' "function in the same way" is a matter of whether they are connected up to the same "information file" (the file with "banks are nice places to walk along rivers" or the one with "banks are institutions that make huge profits by risking taxpayer's money").

[4] Again, this formulation presupposes that all my tokens of 'dogs bark' are synonymous; if you want to avoid any such presupposition, you can put the question as: How do I know that 'dogs bark' as used inside the content marks later in this sentence has the content <<dogs bark>>, and that 'dogs' as used inside the content marks later in this sentence has the content <<dogs>>?

(C) ‘e’ (as I use it) has the content $^{<<}$e$^{>>}$.

By accepting the schema I mean that we believe all instances in which an expression we understand is substituted for both occurrences of ‘e’, and that we are committed to believe similar instances involving expressions we later come to understand. (Some qualifications are needed to handle ambiguity, indexicals, demonstratives and the like, but it’s hard to believe that they couldn’t be given.) The fact that we acquire the body of beliefs about content given by the schema is crucial to the notion of content serving the purposes it serves. That these beliefs are “built into the meaning of ‘content’” in this sense is all we need to legitimate the practice of giving default status to all instances of (C). That’s what justified belief in the instances of (C) consists in; only an absurdly foundationalist epistemology would require more than this.

I think this is a perfectly good answer to the question of how we are justified in believing that ‘dogs’ has the content $^{<<}$dogs$^{>>}$, and maybe to the question of how we know that.[5] (I take the

[5] Attention to indexicality would require a complication in the schema (C): we can no longer have strict homophony. That is, the conceptual truth governing the content of ‘That sound is occurring somewhere far away at this precise moment’ isn’t strictly an instance of (C), it is rather something like

‘That sound is occurring somewhere far away at this precise moment’ as uttered on a given occasion has the content that a certain sound is occurring far from the speaker at the precise time of the utterance.

I take it that such minor transformations from homophony are part of basic linguistic competence, and that the need for them does not alter the basic story, which is that content

notion of knowledge to be sufficiently flexible that it is unclear what more than justified belief is required.)

Nonetheless, there is something important about the epistemic status of our belief that 'dogs' has the content <<dogs>> that this answer doesn't explain: it doesn't explain that this belief is empirically indefeasible (or at least, that it's about as empirically indefeasible as anything ever is). Suppose someone were to say:

> I have found evidence that 'dogs' doesn't have the content <<dogs>>, it has the very different content <<giraffes>>. It isn't that 'dogs' is ambiguous, and on some uses is synonymous with 'giraffes'. No, on *every* use, 'dogs' fails to have the content <<dogs>> but does have the content <<giraffes>>.

This would seem to indicate gross conceptual confusion; it seems impossible to imagine any sensible theory that could make sense of this. But the fact that our learning to use 'content' involves coming to accept all instances of schema (C) isn't enough to explain this.

Consider for instance how we learned to use 'temperature', or (better for my purposes) the comparative predicate 'has higher temperature than'. I take it that a few hundred years ago, part of the ordinary mastery of that predicate was the acceptance of such beliefs as that if one body felt substantially warmer than another to a normal observer in normal conditions then its temperature was higher. Such beliefs would have been taken to be central to the explanatory purposes of the notion of having a higher temperature. These beliefs were in a perfectly good sense built into the content of 'has a higher temperature', and had a default status in that no one was required to argue for them on empirical grounds. They were justified; only an absurdly foundationalist epistemology would deny that.

---

claims for our own language have a very special epistemological status.

Nonetheless, these beliefs were rationally revisable, as is shown by the fact that they were rationally revised: once people started exploring the physical underpinnings of temperature and of feelings of warmth, it was discovered that the beliefs were false, in that things other than temperature-differences (for instance, differences in thermal conductivity) play a substantial role in determining feelings of comparative warmth. (Other things that were probably built into the notion of temperature, e.g. that it was the density of a fluid called 'heat', shared a similar fate.)

But in the case of 'has the content that' it seems impossible to imagine the analogous thing: to imagine discovering that 'dogs' never has the content $^{<<}$dogs$^{>>}$ but instead has the content $^{<<}$giraffes$^{>>}$. So we need a new answer to the question of how the claim that 'dogs' has the content $^{<<}$dogs$^{>>}$ can have the special epistemic status that it has.

I think the answer is that the content marks notation (and its special case, 'that' clauses) is not sufficiently revealing: "content mark terms" like '$^{<<}$dogs$^{>>}$' involve a reference to our language, they are short for something like "the content that 'dogs' actually has today in our language".[6] So

> In our language today, 'dogs' has the content $^{<<}$dogs$^{>>}$ (and 'dogs bark' has the content that dogs bark)

just amounts to something like

> In our language today, 'dogs' has the same content that 'dogs' actually has in our language today (and 'dogs bark' has the same content that 'dogs bark'

---

[6] This of course is very much in the spirit of Carnap 1956 and Davidson 1968. I think the following elaboration of the idea escapes many of the objections that have been raised against Carnap and Davidson, such as those in Speaks 2014.

actually has in our language today).

(The point of the 'actually' is of course to accommodate the fact that if our linguistic usage had been different, 'dogs' *wouldn't* have meant (had the content) $^{<<}$dogs$^{>>}$, i.e. wouldn't have meant what we *actually* mean by 'dogs'.) A formalization that literally quantifies over contents would represent this as

For all linguistic contents c, ['dogs' has c in our language today if and only if actually ('dogs' has c in our language today)];

which is a conceptual truth (or, a logical truth in the logic of 'actually'). Perhaps there is a way to represent the claim without quantifying over contents, but I think that any adequate representation would make it some sort of conceptual or logical truth.

In short: once we understand the logical form of homophonic content attributions to expressions of our own language, we see why knowledge of them can be accounted for without recourse to an incorrigible faculty of introspecting their contents; simple logical knowledge is all that's required.

Of course, this account won't similarly trivialize *non-homophonic* content attributions for terms in our language, like

'bachelor' in our language today means (approximately) $^{<<}$unmarried adult male$^{>>}$;

the account only tells us that this amounts to something like

In our language today, 'bachelor' has (more or less) the same content that 'unmarried adult male' actually has in our language today.

Knowledge of synonymy in addition to logical knowledge is required. Similarly, the account won't trivialize content attributions to expressions of other languages, but merely reduce knowledge of them to synonymy claims (in this case interlinguistic synonymy, as opposed to intralinguistic as in the case of 'bachelor' and 'unmarried adult male'). But all this is as it should be.

Given this, perhaps a more felicitous representation of $^{<<}$e$^{>>}$ (where again, 'e' is to be replaced by an expression of our language) is @LC('e'): @ for 'actual', 'LC' for 'linguistic content'. Because of the quotation marks, this has the advantage of making explicit the linguistic nature of the attribution. (Similarly for intentional content, if one allows that that might be different: @IC('e') instead of $^{<}$e$^{>}$.)

Besides the reason just given for the more explicitly linguistic notation, there is a related one. Someone could have some initial understanding of 'bachelor' without knowing that it means the same as 'unmarried adult male' (if it does, as I'll assume);[7] in double-bracket notation, she starts out knowing that 'bachelor' has the content $^{<<}$bachelor$^{>>}$ but not that 'bachelor' has the content $^{<<}$unmarried male$^{>>}$. But this might seem paradoxical since $^{<<}$bachelor$^{>>}$ = $^{<<}$unmarried male$^{>>}$. Of course it isn't really any more paradoxical than that someone who knows the law of identity needn't know that Hesperus is Phosphorus, but it does show that the "mode of presentation" of the term '$^{<<}$unmarried male$^{>>}$' is important, and the alternative notation @LC('unmarried male') has the virtue of making the linguistic nature of the mode of representation explicit.

There is, to be sure, a worry about making the linguistic item appear explicitly in the representation of content attributions: Church's translation argument (Church 1950), or a more

---

[7] By an initial understanding of it I mean a minimal competence in its use. (This is in opposition to a view that takes understanding as "knowledge of content", where that is construed as knowledge that it has the content .... Of course, by the triviality of the content-that schema, minimal competence in the use of 'bachelor' and in the use of 'content' is enough for knowledge *that 'bachelor' has the content* $^{<<}$*bachelor*$^{>>}$; but this knowledge is totally trivial and has no role in explaining understanding.)

pointed version of that given in Schiffer 1987 (pp. 133ff.). The worry arises from the fact that what's trivial in German isn't

'Hund' hat der Inhalt @LC('dog')

but rather

'Hund' hat der Inhalt @LC('Hund');

and yet on standards of translation that require reference-preservation of the parts, it is the former rather than the latter that translates

'Dog' has the content @LC('dog').

(Similarly for content-attributions for sentences.) Basically I think that all this shows is that those are not the proper standards of translation in this case (any more than they are for the translation of dialog in a historical account of, say, the debates in the Kennedy administration during the Cuban missile crisis). The point of using the 'has the content that' locution in describing a sentence S is to provide an exemplar of the content of S, an exemplar that will be understood by the listener even if S isn't; any reasonable account of "proper translation" (or of "what is literally believed") must accord with this.

For a non-linguistic analogy, suppose that a witness before the Warren Commission described the impact of the decisive bullet by pointing at the place on his own head "where the bullet hit", i.e. analogous to the place on Kennedy's head where it hit. And suppose that in some future investigation someone is asked to give a literal account of the Warren Commission testimony; she will do so by pointing to a spot on her own head, not by digging up the original witness and pointing to the spot on his head. (As it stands there is a disanalogy: her digging up the original witness, or his head, would merely be impractical. We could get a close analogy with a fanciful story on which there are different communities that are by and large invisible to each other, but where with some training a "translator" in one community can learn to see the

members of a specific other community. The Warren Committee witness and the woman asked to report that witness's testimony belong to different communities, but the woman is a "translator" of the other community, i.e. has the special training required to see the witness. In reporting the witness's testimony to other members of her community without this ability, it would not be merely impractical to drag the other witness before her, it would be useless: pointing to her own head is the only way to convey the witness's testimony to them.)

We might call what I've been arguing for *the linguistic exemplar view of standard content attributions*. (As remarked earlier in a footnote, it is very much in the spirit of Carnap and Davidson.) It doesn't say anything about what linguistic and intentional contents *are* (or even, whether literally speaking there are such things); it is merely an account of our standard way of ascribing contents, via 'that'-clauses and by analogous devices for sub-sentential expressions (which are conveniently generalized in the bracket and double-bracket notations). If someone provides a theoretical account of what contents are, then one could attribute contents via the theoretical descriptions of them provided by such an account; but our *ordinary* attributions of content go by linguistic exemplars.

2. Truth conditions. It is natural to think that an important aspect of the linguistic contents of sentences and the intentional contents of mental states of believing, hoping, imagining and so forth is truth conditions. And that an important aspect of the linguistic contents of subsentential expressions, and perhaps the intentional contents of certain "features" or "components" of states of believing etc., is their contributions to the truth conditions of the sentences or states in which they figure.

It's worth noting that 'truth conditions' is somewhat ambiguous. Do 'Hesperus is bright' and 'Phosphorus is bright' have the same or different truth conditions? Uncontroversially, the first has the truth conditions that Hesperus is bright and the second has the truth conditions that

Phosphorus is bright; but are these truth conditions the same or different? On most accounts of *contents* the contents are different (*especially* in communities where the identity of Hesperus and Phosphorus is unknown); but it is common to individuate truth conditions more coarsely than contents (which is why I said that it's natural to view truth conditions as *an aspect of* content). Presumably the idea behind this is that truth conditions should be conceived in terms of possibilities, on a conception of possibility on which true identity sentences between proper names count as necessary.

On either way of construing truth conditions, there is a serious question about the informativeness of the claim that truth conditions (or contributions to truth conditions) are an important aspect of content. This point doesn't depend on which way we construe truth conditions, but since it is perhaps less obvious on the coarse grained or possibility conception, I will primarily work with that.

First some background. A large part of our grasp of the notion of truth consists of our acceptance of the instances of the schema

(T) For any sentence (or sentence-token) S of any language, if S has the content that p, then S is true iff p.

'p' is schematic, for sentences of the meta-language—here, English. (More generally, we could allow it to be for *fully parameterized formulas* of the meta-language: formulas together with assignments of objects to their free variables. In effect, this generalizes (T) to

(T_{gen}) For any sentence (or sentence-token) S of any language, if there are objects $o_1,...,o_k$ such that S has the content that $F(o_1,...,o_k)$, then S is true iff $F(o_1,...,o_k)$.[8]

---

[8] The schemas (T) and (T_{gen}) are indefinitely extensible: we not only accept instances for sentences p and formulas F in our language today, we have a commitment to instances for sentences and formulas we come to understand in the future. But it isn't entirely clear that this is

This is important, but for simplicity I'll work just with (T) in what follows.)

If our interest includes modality we can also generalize (T), to

(T^+) For any sentence (or sentence-token) S of any language, if S has the content that p, then for all worlds w, S is true at w iff in w, p;

where the notion of world is to be understood in accordance with the kind of modality in question.[9] (And we can make an analogous modal generalization of (T_{gen}).)

Of course if we aren't working in a formalized language but in a language like English that contains ambiguous expressions, we must take the allowable instances of (T) and (T^+) to be only those in which the substituend for 'p' is disambiguated in the same way in each occurrence. But that is nothing special about (T) and (T^+), it is standard in the interpretation of schemas: logical laws in schematic form, such as

($\wedge$E) If p and q then p,

---

relevant in the present context, since we don't yet understand those instances. (More on this in Section 5.)

[9] (T) is part of our ordinary understanding of truth, and the same is true of (T^+) to the extent that our ordinary understanding can be said to include the notion of possible world. Because of the semantic paradoxes, there *may* be reason to revise this ordinary understanding by restricting (T) and (T^+), e.g. to sentences S that don't contain semantic terms or some less drastic restriction. My own view is that this isn't necessary, that the best way to avoid the paradoxes is by restrictions on the logic, but this is not the place to discuss that.

would obviously fail otherwise. In addition to ambiguities (both lexical and syntactic), English contains indexicals, demonstratives and so forth, but just as with (/\E) we demand that the interpretation of indexical and demonstrative elements in the substituends of 'p' not vary from one occurrence to the other.

Presumably we not only accept the instances of (T) and (T^+), but accept that they have some kind of special status. Crudely put, instances of each are both conceptual truths and necessary truths, more or less like the instances of (/\E).[10]

The reason for doubting the informativeness of the claim that truth conditions are a component of the content of sentences is that by the linguistic theory of content-attributions, (T)

---

[10] In calling them conceptual truths I'm not claiming that their truth is explained by the meaning of 'true' and/or 'has the content that' (and/or 'if' and 'iff'): I'm not sure what that would even mean. I'm also not claiming that it's inconceivable that we could give them up without change of meaning, but merely something like: it's hard to clearly envision any rational way to give them up, and someone's rejecting them is prima facie evidence of their being conceptually confused. At a minimum, "conceptual truths" are extremely well-entrenched parts of our theory that we would challenge only under extreme circumstances.

reduces to something like

For any sentence S, if the linguistic content of S is @LC('p') then S is true iff p;

and analogously for (T^+). But when we apply these to a sentence of our own language as we actually use it today, they amounts to just:

(ST) 'p' as we actually use it today is true if and only if p;

(ST^+) For all worlds w, 'p' as we actually use it today is true at w if and only if in w, p.

And so, for the instances of (T) and (T^+) to be conceptual truths, those of (ST) and (ST^+) must be as well. (Similarly for them to be necessary truths, though their necessity is a rather trivial addition since it has simply been built in by use of the actuality operator.) Saying that 'dogs bark' (as we actually use it today) is true (at w) if and only if dogs bark (in w) thus seems to have no empirical content, it is merely a consequence of our notion of truth.

Of course, the following are still empirical claims:

'Dogs bark' as used in 1800 is true (at w) if and only if dogs bark (in w)

'Hunde bellen' as used in Germany today is true (at w) if and only if dogs bark (in w).

But the empirical content of these is no greater than the empirical content of synonymy claims:

'Dogs bark' as used in 1800 is synonymous with (has the same content as) 'dogs bark' as actually used today;

'Hunde bellen' as used in Germany today is synonymous with 'dogs bark' as actually used today.

Given this, it wouldn't seem very informative to say that truth conditions are an important part of the content of sentences in our actual current language. And the analogous claim for other

languages wouldn't seem informative either unless it could be argued that truth conditions play an important part in the notion of synonymy between other languages (including other time-slices of English) and our actual current language.

And how *could* truth conditions play an important role in the latter synonymy relation? The idea would have to be that sameness of truth conditions between the other language and our (actual current) language played such a role. But that seems impossible if claims about truth conditions for our (actual current) language are empirically empty.

Of course, I'm not denying

(S) Sentence tokens of different languages (one of which may be our current one) are synonymous (have the same content) only if they have the same truth conditions.

What I'm denying is that this gives an informative account of synonymy or sameness of content. It could do that only if we had a notion of truth conditions for other languages that was not itself based on sameness of content. The claim (S) needs to be viewed not as partially explaining synonymy in terms of an independent notion of truth conditions, but as partially explaining the notion of truth conditions for other languages in terms of an independent notion of synonymy. The notion of truth-conditions is egocentric, not suitable as the basis for an objective semantic theory.

I've been speaking of the linguistic content of sentences, but what about the intentional content of mental states of believing, hoping, imagining, and so forth? Some of these mental states—states of *explicitly* believing, *explicitly* hoping, etc.—seem intimately related to language, in such a way that any account of truth conditions for the sentences of a language extend fairly directly to these states of a user of the language, without appeal to any synonymy-like notion between those states and sentences.[11] (Or so it seems to me, but nothing will hang on

[11] If you like, that's because *explicit* belief involves the acceptance of sentences; and analogously

this.) If this is so, then in particular the attribution of truth-conditions to ones own explicit mental states (the actual current ones) is as empirically empty as is the attributions of truth conditions to ones own current language. But it is far from obvious that this connection to one's language extends to all of one's inner states; to the extent that it can't, we need a synonymy-like relation between states and sentences, e.g. the relation E that a state bears to a sentence if the intentional content of the state is the same as the linguistic content of the sentence.[12] (That formulation of E requires that intentional contents and linguistic content be the same sort of things; but it is easy to modify the formulation to fit other conceptions, for instance if intentional contents correspond to some sort of equivalence class of linguistic contents. There's no point worrying too much about this at the moment since I'll be casting doubt on the clarity of these synonymy-like notions in the next section.)

So the basic picture is: the notion of truth conditions has a clarity, independent of the notion of synonymy, for the sentences of our current language and perhaps for some of our own belief states, imagination states, and so forth. But its clarity there is due to the fact that attributions of truth conditions in these cases have no empirical content. (Quine 1970 called this the "immanence" of truth.) In cases where the attributions have empirical content, their empirical content consists in the content of some synonymy-like relations to the attributor's own sentences or own states of believing, imagining, etc.. And I'll be going on to argue that these synonymy-like relations are not altogether clear.

---

for explicitly hoping etc..

[12] I'm taking our language as basic because I've been concentrating on *linguistic* attributions of truth conditions. If one wants to concentrate on *mental* attributions of truth conditions, perhaps it is attributions of truth conditions to our explicit beliefs that is basic, and attributions of truth conditions to our sentences dependent on them. I doubt that much hangs on the difference; but attributions to linguistic items are easier to talk about, which is why I focus on them.

3. Synonymy and the Normativity of Meaning. In formulating the linguistic exemplar view of ordinary attributions of content, I've been writing as if there were such things as linguistic and intentional contents, such that two sentence-tokens being synonymous or "having the same content" as we normally understand that involves their standing in the linguistic content relation to the same linguistic content c (and analogously for intentional states and intentional content). This strikes me as highly dubious. Indeed, I don't think that these synonymy notions ("having the same linguistic content" and "having the same intentional content"), as we ordinarily understand them, are equivalence relations. Especially, they aren't equivalence relations *when applied across communities whose members have markedly different overall theories*.

There are many examples in the literature, often involving a community C that splits into two fragments F_1 and F_2 that lose contact with each other. As the members of F_1 and F_2 learn more, they express their revised views using their old vocabulary, but find different ways of doing this, with the result that certain terms in F_1 become clearly non-synonymous with the corresponding term in F_2: were the fragments to start interacting again they would never translate each other homophonically. And yet neither of the offshoot communities recognizes a change in their own usage from before the fragmentation: each would translate C homophonically. This scenario or something similar to it was used in Mark Wilson 1982 (e.g. the example based on *The Island of Lost Women* that opens the article), and in Mark Lance and John Hawthorne 1997 (the modified Salem witch story, pp. 45 ff). Another example, involving the term 'mass' as used in Newtonian mechanics, was implicit in Field 1973—though that paper attempted to downgrade its significance in a way that I now think ill-advised.

Field 2009 had another example, involving logical vocabulary, in particular negation as used in three different logics. Consider three isolated logical communities. Community I is composed of classical logicians. Community II is composed of hard-core intuitionists: people

who think that intuitionist logic is the correct logic, so that classical logic is simply incorrect; in particular, incorrect about negation. Community III is composed of people who weaken classical logic in order to keep the truth schema unrestricted without paradox; since the paradoxes arise for intuitionist logic too, this requires restricting some intuitionistic laws, but it doesn't require all of the intuitionistic restrictions on classical logic. Consider now how the negation sign is to be translated. I think it's pretty clear that the members of Communities I and II should translate the negation sign of each other homophonically—no other obvious alternative is available. But an advocate of typical logics of paradox in which 'not A' is distinguished from

'if A then $\perp$" should translate the classical logician's negation by the former and the intuitionist's

by the latter.

One might perhaps say that in cases like these, the terms in one community strictly differ in content from the terms in the other, where strict sameness of content is an equivalence relation; they merely approximate each other in content, and no one should expect approximate sameness of content to be an equivalence relation. Perhaps this is right, but if so then it is the notion of approximate synonymy that is important to us in our attributions of content to foreign terms. (What we're concerned with at the moment is ordinary content attributions; later in the paper I'll turn to the possibility that there are more theoretical roles that contents might play, and in that context the claim that the synonymy is only approximate may have more relevance.)

In the above examples and in others, the intransitivity (and possibly asymmetry) of the synonymy stems from the fact that we make content attributions as aids in communicating, and that this communication serves a variety of important purposes for us: it enables us to learn from

and teach others, it enables us to be influenced by and to influence others, and it enables us to engage with others in many other kinds of social interactions. A good translation is one that is useful for these purposes. The purposes may be somewhat in competition, and no translation will fill even a single purpose perfectly; this combination of facts gives rise to significant indeterminacy in what counts as a good translation, and also underlies the intransitivity and/or asymmetry.

Given this, our notion of truth conditions applied to sentences in another community amounts to something like this:

(T^{+}_{eval}) For any sentence S of another linguistic community, if S is well-translated as 'p' as we currently actually use it, then for all worlds w, S is true at w iff in w, p.

This makes explicit that our notion of truth-conditions as applied to another community is an evaluative notion, it is a notion dependent on the idea of goodness in translation. We simply don't have any translation-independent sense what it is for the Salem-ites claim that Sarah was a witch to be true; it is true relative to some translations but not relative to others. Similarly for the representations of the frog who reacts to a moving black blob that is not an insect.

I've been speaking mostly about synonymy claims between distinct linguistic communities, and the application of our notion of truth to members of linguistic communities other than ours. But what about synonymy claims between members of the same community, and truth attributions to other members of our community? The only real difference between the intra-community case and the inter-community case is that in the intra-community case the social interactions relevant to good translation are far broader. One consequence of that difference is that even central disagreement within a linguistic community about the "core theory" associated with a term is far from decisive as a reason not to translate homophonically.[13]

---

[13] In the splitting-community examples above, the lack of interaction between the communities

The point is familiar, but an especially forceful case is made in Williamson 2007, pp. 85-98. For instance, Williamson correctly stresses that we don't normally regard disagreement over central principles such as Modus Ponens (such as that between McGee 1985 and Williamson) to signal a difference of meaning in 'if...then', and that it would tend to disrupt communication to so regard them: the homophonic translation is to be preferred despite the central difference in principles. (And if it's appropriate to translate homophonically then attributions of truth-conditions should proceed homophonically.)

But I think it would be an error to draw from Williamson's discussion the moral that there is a completely clear notion of sameness of intra-community meaning, such that people who disagree about whether Modus Ponens must be restricted determinately mean the same by 'if ... then'. One way to see that this would be an error is to imagine that McGee started to have uncertainty about his conclusion that Modus Ponens needs restriction, and began working with two distinct analogs of the indicative conditional, one obeying Modus Ponens without restriction and one not; let's suppose that he devotes considerable effort to studying the relationships between the two, so in his private work he employs distinct subscripts on 'if' to keep them straight. But in ordinary communication with others who aren't up on his work he drops the subscripts. The two subscripted notions clearly differ in meaning for him, and his unsubscripted uses clearly don't have a determinate meaning distinct from those two, so it would be inappropriate to say without qualification that McGee's unsubscripted 'if...then' is synonymous with Williamson's.

---

prior to translation minimized the salience of factors in translation other than "core theory". But even in those cases, such factors aren't entirely absent, a fact which increases the degree of indeterminacy in the translations. I don't think that this effects my main point, which was about whether we have reason to regard interlinguistic synonymy as an equivalence relation.

I conclude that the interpretation even of the utterances of members of one's community is a matter of good translation (good for the purposes at hand); it's just that in the intra-community case there are considerations that give extra weight to the homophonic translation.

What about synonymy between one's earlier uses of a term and one's present uses (when the term is unambiguous in both prior language and present language)? At least when the time-difference isn't large, there are still stronger pragmatic reasons to privilege homophonic translation (and hence homophonic truth conditions), barring clear reasons to the contrary, despite even fundamental change in theory. But here too there are situations where this isn't so clear: e.g. cases where the change of theory was prompted by a discovery, and it was somewhat arbitrary how we decided to use our words given the discovery. (Such cases, e.g. 'mass' as used by Einstein before and after the discovery of special relativity, are single-person analogs of the splitting community cases above.) Translation of (and attributions of truth conditions to) one's earlier utterances, like translation of other communities, is evaluative.

These remarks are largely in agreement with Lance and Hawthorne 1997. They call interpersonal synonymy and truth-conditions *normative* notions rather than *evaluative* notions, and perhaps that's better. The change is not huge, it is from 'good' to 'ought':

(T^{+}_{norm}) For any sentence S of another community, if one ought to translate S as 'p' as we currently actually use it, then for all worlds w, S is true at w iff in w, p.

It is natural to regard (T^{+}_{norm}) as having broader application than (T^{+}_{eval}): for it's natural to say that because of our interest in communication, we ought to try to translate other people as best we can, even if no such translation is really all that good. I think the broader application is probably an advantage to (T^{+}_{norm}) (though it's possible to read 'well-translated' in (T^{+}_{eval}) in a loose way on which there is no difference). There may also be a disadvantage to the move from evaluative to normative talk: I think that it is more natural to read (T^{+}_{norm}) as requiring some kind

of hyper-realist view of 'ought's than it is to read (T^{+}_{eval}) as requiring a hyper-realist view of goodness. But I'm happy to go with (T^{+}_{norm}), on the stipulation that it is neutral as to the status of the 'ought's: that it leave open, for instance, that they be read in some kind of broadly expressivist manner.

Lance and Hawthorne describe their view as one in which meaning (content) is normative, and I'm not entirely happy with putting my own view that way. A very minor reservation is that the present discussion is concerned only with ordinary content attributions, and I haven't *ruled out* that talk of content divorced from such attributions might serve some explanatory purposes. But I will not in the end attach much importance to this. A more significant reservation is that the examples we've discussed have concerned only interpersonal synonymy (taken in a broad sense that includes synonymy between temporally distant time-slices of the same person), and the normativity (or evaluativeness) of synonymy plays far less role in the intrapersonal (and broadly intratemporal) case.

For I think that there's a lot to Quine's suggestion in sec. 11 of his 1960 that a passable conception of *intrapersonal* (and broadly intratemporal) sameness (and similarity of) content can be founded on a notion of sameness (and similarity) of cognitive or epistemic role. (This of course presupposes that *that* notion is independent of synonymy; I won't defend that presupposition here, but think it correct.) Quine developed the idea in a crudely behavioristic way which limited its scope, but the guiding idea is broader. The idea is that 'Hesperus is bright' differs in content for a person from 'Phosphorus is bright' because different things could count as evidence for one than for the other; and that 'Hesperus' then differs in content from 'Phosphorus' because substitution of one for the other in a sentential context (e.g. in '... is bright') changes what could count as evidence for the resulting sentence. (Of course that criterion would be hopeless for interpersonal synonymy, but as he'd said earlier, "it is an open

question how satisfactorily [interpersonal synonymy and intrapersonal synonymy] can be subsumed under a single general ... synonymy concept" (Quine 1953, p. 56).) As Quine 1960 notes, it is somewhat unclear whether this test allows for a difference in content for a subject who is thoroughly convinced in the claim 'Hesperus is Phosphorus'; if we restrict to fairly local evidence it doesn't, but if we allow for evidence sufficiently extensive to undermine that total confidence it does. Of course it may be difficult to draw a line between extensive evidence that "keeps the contents fixed" and extensive evidence that motivates "change of content". But that's only objectionable if one's goal is a super-precise notion that cleanly separates content from theory; if like Quine we reject that as an illusory ideal, but still want to say something to illuminate our intrapersonal synonymy judgements, I think this epistemic account admirable.

The point is that whereas Frege thought that one needed a difference in sense to *explain* how evidence for one can fail to be evidence for the other (for the same person at the same time), Quine's view rejects the explanation and takes the difference in content to be *constituted by* the fact that what we count as evidence for one fails to be the same as what we count as evidence for the other. (Perhaps one could broaden the account to allow for differences in the psychological route by which evidence counts for the sentences; I don't think that would fundamentally alter the spirit of the account.)

I wouldn't want to say that this account of intrapersonal synonymy is thoroughly non-normative, because there are probably normative elements in any reasonable elaboration of "what counts as evidence". Still, these normative elements are quite different from those that go into interpersonal synonymy judgements (especially those between distinct linguistic communities).

Though the normative elements that go into interpersonal synonymy judgements are absent *from the account of intrapersonal synonymy above*, I don't think they are entirely absent

from ordinary judgements of intrapersonal synonymy. For (as I think Quine also pretty much noted in the above section of his 1960) we can build a speaker's relations to his or her linguistic community into intrapersonal synonymy. That is, we can derivatively regard two expressions as differing in content for a speaker if there are competent members of their linguistic community for which the expressions differ in content in the non-derivative sense provided by the evidential account. It isn't obvious to what extent a conviction that 'Hesperus' and 'Phosphorus' differ in content for us today depends on *including in our linguistic community* those who don't firmly believe 'Hesperus is identical to Phosphorus', as opposed to relying on evidence that might undermine the identity even for oneself (and/or relying on considerations of *evidential route*). There's no need to decide, unless there's a reason to make intrapersonal content a precise notion; and I doubt that there is any such reason.

In short, the slogan that meaning or content is normative may be misleading in failing to distinguish intrapersonal from interpersonal sameness of content; while the former as well as the latter may have normative elements, the normative elements that are central to the latter are far less central to the former.

But this is really something of a digression, because it is only interpersonal synonymy that plays a role in truth conditions, given that truth conditions for our own sentences are automatic from (T^+) together with (C).

4. Pleonastic propositions. I have mostly avoided the use of the term 'proposition', not because I see no possibility of legitimate use for it but because I think its use tends to obscure important facts, including but not limited to some of the facts I've been arguing for. Even talk of *pleonastic* propositions is not always employed innocently.

In Schiffer 2004, pleonastic propositions are introduced by the idea that "The proposition that dogs bark is true" is a harmless variant of "Dogs bark", and more generally that an instance

of the schema

The proposition that p is true

is a harmless variant of the corresponding p. (Schiffer makes an exception for paradoxical instances, though if one is willing to restrict classical logic a bit this isn't really necessary.) I'm fine with this. It is common to point to an analogy that Frege used in a different context: the analogy to the direction of lines. Unless one wants to indulge in ontological scruples which aren't at issue here, the claim that two arrows are pointing in the same direction just seems a harmless paraphrase of the claim that they are parallel.

This is all fine. The danger comes when one assumes that propositions play a role not supported by their pleonastic introduction. This could happen even for directions. Imagine someone arguing as follows:

> Consider two arrows, perhaps far away from each other. Each is straight, so on the pleonastic notion of direction, each has a direction. If these directions are the same, the arrows are objectively parallel; if the directions are different, the arrows are objectively non-parallel (though if the directions are close to each other they are close to parallel). So it is an objective question whether the arrows are parallel. But if space were non-Euclidean, there would be no objective parallelism between the arrows; parallelism would be relative to a path of transport. This is an a priori proof of the Euclideanness of physical space!

This argument, I hope we'll agree, is absurd. The reason is that if at a given point of space we introduce pleonastic directions, they are directions that automatically apply to directed lines *only at that point of space*. It isn't that that point of space is metaphysically privileged (unique among others in that there directed lines have objective direction!), it's simply that directions *introduced there* don't automatically apply at other points. (On fairly minimal assumptions they

do apply elsewhere *relative to a path of transport*. And we might sometimes have reason to single out some paths of transport as better than others: for instance, if the arrows are near to each other, we might want to consider paths that stay near to both and don't loop around a lot, and with such a restriction, parallelism relative to one allowed path won't differ *by very much* from parallelism relative to another allowed path.) At another point of space we can also introduce directions; but the directions introduced there can only be compared to the directions introduced at the first point relative to a path of transport. Directions, in short, are *local entities*. If space is Euclidean there is a path-independent way to correlate these local directions, i.e. we can introduce global directions; but the Euclideanness of space is a substantive empirical assumption, and without it, global parallelism makes no sense.

Similarly, we can imagine someone arguing as follows:

> Consider two utterances, made by different people; or two belief states of different people. Each expresses a (pleonastic) proposition, i.e. has such a proposition as its content. If these propositions are the same, the sentences or belief states are objectively synonymous or the same in content; otherwise, they objectively differ in content (though perhaps their objective content is similar). So sameness of content is totally objective. This is an a priori proof that what was said about synonymy in Section 3 is incorrect!

This argument is equally absurd, and the reply is exactly analogous to that in the case of directions. If for a given person and time we introduce pleonastic propositions, they are propositions that automatically apply to sentences or belief states *only for that person*. It isn't that that person is metaphysically privileged (unique among others in that for him or her, sentences and belief states have objective content!), it's simply that propositions *introduced there* don't automatically apply to the states of other people. (On fairly minimal assumptions

they do apply to other people relative to a method of translation; for instance, relative to a homophonic translation or a conventionally employed translation or a translation that is useful for certain purposes. We may be able to constrain the allowable translations sufficiently so that barring major overall differences between the people, sameness of content by one allowable translation and sameness of content by another won't differ *by very much*.) For another person we can also introduce pleonastic propositions; but the propositions introduced there can only be compared to the propositions introduced for the first person relative to a translation. Pleonastic propositions are local entities. Of course someone might make assumptions on which there is a translation-independent way to objectively correlate these local propositions, i.e. to introduce global propositions; but substantive empirical assumptions are needed for this to make sense. The assumption of *global* propositions is in no way justified simply on the basis of the pleonastic theory.

5. Untranslatable utterances and states. The notion of translation is, as I've said, a flexible one: we tend to make sense of other people by translating them, even though we recognize that there may be important ways in which our translation doesn't capture the subtleties of their views: if their views are significantly different from ours, no translation can capture all the subtleties. Since it's important to make sense of other people, it's important to translate even given the limitations. Sometimes a particular term of theirs presents a special difficulty for translation; in that case, it is common to sometimes leave it untranslated (with informal directions as to its use), leaving the reader to incorporate it into his or her idiolect. Those who work on Aristotle or Kant are particularly prone to doing this.

Another example of incorporation is with the use of proper names. I hear Mary use the name 'George', and take her to be talking about someone other than George Soros, or either of the George Bush's, or George Washington, or any other George I'm familiar with. So I

incorporate her word ‘George’ into my idiolect (with an implicit subscript to keep information about this new George separate from my information about the other Georges), and learn to use it by listening to what Mary says (not necessarily accepting her claims at face value).

It’s worth remarking that my doing so doesn’t seem to rely on any antecedent grasp of the notion of reference. Of course if I start out with a notion of reference governed by the schema

For any object o, ‘t’ as I currently actually use it refers to o iff o is identical to t,[14]

then I will extend this to include the new use of ‘George’, and will be able to say that this use of ‘George’ refers to George (as currently used). (Unless of course I believe that Mary is lying or delusional, in which case I may say that ‘George’ as she is using it doesn’t refer to anything.) And I will say that the relevant instances of Mary’s term ‘George’ are equivalent to my incorporated term ‘George’, hence refer to this George. So, if I have this trivial notion of reference available, there’s an equivalent to the incorporated name, viz. “the referent of (such and such tokens of) Mary’s term ‘George’”. In other words, as Brandom 1984 says, reference-talk can be used as a device of cross-person anaphora. The same account applies straightforwardly to demonstratives: if Mary says ‘that’ and I don’t “know what she’s talking about”, i.e. there’s no term in my public language or “language of thought” that I am willing to regard as equivalent to hers, then I can either incorporate it, or use the above “disquotational” notion of reference and speak of “the referent of her use of ‘that’ on this occasion”. No use of the notion of reference not fully explained by the schema (understood to extend to incorporated terms) is required for this: Brandom’s anaphoric function for ‘refers’ can be viewed as a

[14] Or more generally:

For all N, if N is to be translated into my current actual language as ‘t’ then for any object o, N refers to o iff o is identical to t.

consequence of the indefinitely extensible reference schema.

What I've said for reference extends to truth. Suppose (to adapt an example from Shapiro 2003) that I'm told that a great set theorist has asserted some sentence S that I can make almost nothing of, because it contains many words that are beyond my understanding. But I don't need to be able to understand it *well* to be able to incorporate it into my language with some minimal role. (As Shapiro emphasizes, even that minimal role may still be important: I may appreciate that the set theorist regards it as an explanatory hypothesis that, together with other set-theoretic claims that I may also only minimally understand, entails consequences in arithmetic that I do understand.) And with even a minimal understanding, my truth schema (T^+) commits me to the relevant instance of the truth schema for my incorporated sentence (let's call it S*). And since I take this incorporated sentence S* to be equivalent to the set-theorist's S, that means that I regard the claim that her S is true as having the (ill-understood-by-me) truth conditions that .... (where into the blanks go my incorporated S*). So I come to regard 'The set-theorist's sentence as true' as simply an alternative means of incorporating the set-theorist's sentence. It may be an especially convenient means of incorporating it, if I find her sentence hard to remember or hard to pronounce. Thus the anaphoric function of 'true' emphasized in Grover, Camp and Belnap 1975 can be viewed as deriving from the truth schema given its indefinite extensibility, in complete analogy to what I've argued for Brandom's anaphoric function of 'refers' and the reference schema.

This claim that the anaphoric functions of 'refers' and 'true' should be seen as deriving from the indefinite extensibility of the schemas isn't really essential to what I'm saying: I'd be perfectly happy to posit that as a separate function of 'refers' and 'true' if I thought it necessary. (I once thought that Grover, Camp and Belnap's theory less "deflationary" than it initially appeared because of the heavy role of 'that' clauses in their formulation, but I now think I was

wrong: the clauses can be read in the same translational spirit I've advocated in the paper, with no dependence on any non-pleonastic or non-local conception of proposition.)

In previous work I occasionally said that our notion of truth should be limited to sentences (and belief states) that could either be translated into our language on the flexible standards above or were understood well enough to be incorporated into our language. Given the flexibility I've just indicated in the relevant sense of translation and incorporation, it isn't clear to me that this is wrong, but it seems unnecessary: it would suffice to say, as I've in effect done here, that we have little understanding of the application of truth to sentences that we think untranslatable and don't understand well enough to usefully incorporate.

6. Representational content in explanation.

I take it as beyond serious doubt that there is *a level* of psychological theorizing that does not require a notion of representational content (that is, of intentional representation talk, e.g. of a state representing that ...). It requires instead talk of computational processing, together with talk of causal relations between the organisms and their environment. E.g. a psychological theory of how a frog catches flies might be an elaboration of the idea that flies trigger the frog's motion detectors, which act in certain complicated ways that result in the frog's tongue moving out in the direction of the fly. We might informally talk of the motion detectors as "representing" flies, but the kind of explanation I'm envisaging is complete on its own without such representational talk. (The envisaged explanation would involve there being a correlation between the state of the motion detector and the state of whatever it is that triggers it, but of course we often talk of correlations when we don't talk of representation: e.g. we don't talk of the movement of the frog's tongue as representing either the flies or the motion detectors.)

Such a detailed model of the frog's visual system might or might not be computational in the stricter sense that it employs something analogous to symbol-manipulation in accordance

with syntax-like rules. If that (rather vague) characterization of the stricter sense doesn't well fit the frog's visual system, it is almost certainly accurate for more complex systems such as deliberative reasoning in humans. Here it is natural to speak of items in the reasoning process as "representations"; still, it is clear that there is *a level* of theorizing about it which would simply employ the "syntactic" laws of how "internal tokens of strings of symbols" interact with each other and straightforward causal laws connecting "peripheral tokens" with the organism's environment; representational content would play no role in such theorizing.[15] Admittedly, an accurate such theory for a given organism would be so extraordinarily complex that it would be way beyond the bounds of practical use: at best it would only be practical to use highly oversimplified theories if they were written in this way. But I take it that few doubt that accurate such theories of a given organism are in principle possible. (And to the extent that a large class of organisms work similarly, such a theory can be made to work for the whole class by moves such as generalization over parameters; though this very likely requires more idealization, i.e. sacrifice of accuracy.)

The reason for making these rather uncontroversial points explicit is to give background for the question of what explanatory roles representational *content* plays, within theoretical psychology and/or within ordinary psychological explanations.

It is clear that much in ordinary psychological explanation accords well with the egocentric picture. If I learn that someone has just seen worrisome x-rays of his daughter's

---

[15] The laws would involve syntactic types, but the notion of type is relative to a level of theorizing; here the relevant notion of type is purely computational, so though representational content is relevant to other notions of type it would not be relevant here. (If the computational laws are intended as laws about a particular organism, then the computational types can be individuated using purely intra-organism relations.)

lungs, I can explain his behavior by imagining myself to have been presented with similar evidence about my own daughter and imagining ways I might behave under those circumstances. (See Gordon 1986 for a development of this which is congenial to the egocentric approach.) But the deepest roles that content plays in psychological explanation have to do with the notions of truth and of truth conditions.

On the view of truth conditions advocated earlier in this paper, the application of our notion of truth to our own sentences is governed by (ST^+) and can seem rather trivial, and its application to other languages is less trivial than that only because of its reliance on translation. But even in the case of its application to our own language, there is a well-known point about why the notion of truth is important: using that notion enables us to formulate generalizations that would not be formulable without that notion or some comparable logical resources. The point was perhaps first made explicit in Quine 1970, though it is certainly implicit in Tarski 1956. In a language without a device of quantification into the predicate position, and no "immanent" truth predicate of the sort described in Section 2, there is no way to generalize over the instances of, say, the schema of mathematical induction; but with such a truth predicate I can, by saying that all instances of the induction schema are true. Since the claim that an instance is true is conceptually equivalent to the instance (taking 'conceptual equivalence' to mean at minimum, that the biconditional is an uncontested part of our theory), the generalization serves the purposes of a universal quantification into predicate position.

Truth as a device of generalization plays an especially important role in our theories of ourselves and others: it enables us to state *reliability generalizations*, generalizations about the circumstances under which we are likely to have true beliefs (as expressed in our own language), and the circumstances in which we aren't likely to.[16] For instance, the following are true of me:

---

[16] There are other important ways in which 'true' is used as a device of generalization in

- If there's a large piece of fancy furniture in the room I'm in, then under normal circumstances it's highly likely that I'll have a true belief not only that it's there but what type of furniture it is (chair, couch, table etc.); but I'm extremely unlikely to have a belief as to the designer, and if for some reason I do it is unlikely to be true. Similarly, I'm likely to have a *rough* belief as to its size, and the belief is likely to be true; but I'm unlikely to have a belief as to its length in millimeters (barring special circumstances not mentioned here), and if I do that's very unlikely to be true.
- If there's a medium sized domestic animal 50 yards in front of me, then if and only if I'm wearing my glasses am I likely to have a true belief of its presence and of its type.
- If I hear a frequently-played classic rock song from the 1960's or 1970's I'm pretty likely to have a true belief as to the singer and name of the song and many of the words; whereas if I hear a random quote from Trollope I'm unlikely to have a true belief as to the novel from which it came.

Different reliability generalizations are true of different people, though of course there are some that are true of virtually everyone, others true of virtually all farmers, others of all lawyers, others of all educated Americans, and so forth. And it would be hard to overestimate the importance of such generalizations in interacting with people, and in predicting and explaining

---

explanations. For instance, I can give an incomplete but still contentful explanation of your symptoms by saying "I bet the same thing is true of you as was true of me last week"; if I don't know how to explain my own malady, this may be the best I can do. Because of our ignorance of the workings of the mind, this use of 'true' is common in psychological explanations, but has nothing particular to do with representation.

both their behavior and the extent to which they successfully carry out their plans. A foreign minister to France who did not have reliable information about the names of the leading politicians of Europe would predictably fail in his job. (I can say this without myself knowing the names of those politicians, so the notion of reliability (and hence indirectly, the notion of truth) is playing a crucial role.) A pilot who could not reliably tell the approximate airspeed of her plane when flying without visual cues would be in serious trouble. (I could say this even if I didn't know how it is that she bases her actions on her beliefs about the airspeed.)

It's pretty clear in outline how we could in principle dispense with content attribution in explanations of an individual agent, if we knew enough: e.g. in explaining how the pilot carries out a simple task like keeping the airspeed in a safe range at constant power, all we really need is a parameterized class of belief states such that (i) the value of the parameter of her belief state is correlated with the airspeed, and (ii) if she is in a state with high parameter she pulls the yoke back and if she is in one with low parameter she pushes it forward. As you complicate the pilot's task you need to complicate this, though without losing the basic idea. Still, once it gets much more complicated, talk of representational content is a practical necessity. And when generalizing over different agents, it takes on additional roles, since it is needed to abstract common patterns that are implemented in the different agents in varying ways.

The reliability generalizations two paragraphs back were made using the notion of 'believes that' or some related notion, so on the view presented in this paper, their application to others relies on a notion of good translation. The reliance on translation isn't very evident in these particular examples, where standards of translation are likely to be clear. In other cases, for instance in describing people with deep theoretical differences from us, translation is less objective, and because of this we need to state reliability generalizations in a more nuanced way. The ancient Greek who often used a phrase that might be translated as 'Zeus is throwing

thunderbolts' might be unreliable relative to this translation but reliable relative to a translation as 'It's thundering'; the reliability relative to the latter translation is important in explaining his ability to keep safe in a storm, while the unreliability relative to the former is important in explaining the futility of his attempts to control the weather. Something similar can be said about the dog who we'd be inclined to describe as believing that his master has come home, a belief that presumably doesn't track recent changes in legal ownership. The "exact content" of the belief, if sense can be made of this, isn't of huge explanatory relevance. The content attribution serves rather as a rough and ready way to allude the important features of the belief state.

There is room to dispute the extent to which good translation fails to be objective, but whatever one's view on that, reliability generalizations like those above are extremely important. For this reason it is totally misguided to object to a view of content based on a trivial disquotation schema for our own language or idiolect, plus translation, that it somehow "cuts language off from the world". Important connections between language and the world hold in our own case, despite the triviality of the content schema there. Among these important connections, reliability generalizations are preeminent; though in the explanation of the reliability generalizations, various other sorts of factors that non-translational theorists of content have emphasized play an important role. For instance, such reliability as I have in my beliefs involving the term 'Napoleon' is due to a network of transmission of beliefs, in which I have acquired beliefs involving the name from other users, and they have acquired such beliefs from still others, and so on back to those with direct observational access to Napoleon and his deeds; this causal network has multiple independent chains, and contains experts that have investigated these chains systematically, so that the chance of large errors surviving among experts in the chain isn't that high. In addition to such causal relations to the external world and to fellow language users, internal factors play an important role in explaining reliability generalizations;

for instance, the reliability of my logical beliefs is explained in part by my having internalized inferential procedures for the various logical connectives (and in another part by my recognition of the inferential procedures I've internalized). In short, factors in content which non-translational theorists have emphasized, such as inferential role, causal networks of transmission, and the role of experts, all have a place in explanations on the egocentric or translational picture, even if they are not put directly into the account of content. And because of their importance in explanations, it is factors like these that are primarily taken into account when translating.

One could try to build an account of content directly out of such factors as inferential role, causal relations to the environment, and reliability relations; and perhaps a sophisticated enough attempt at so doing would yield explanations no different from what is delivered by a development of the more egocentric approach I've advocated. But I think the egocentric approach more direct and less likely to lead to pointless verbal debates over which factors are "really part of content". We can all agree that owls who more reliably predict which way their prey will turn are, other things being equal, more likely to survive than their less reliable fellows; the details of our "translation" of the owl's state, and of our decision as to its "exact content", doesn't seem very important to the explanation.[17]

REFERENCES

Brandom, Robert 1984. "Reference Explained Away". *Journal of Philosophy* 81: 469-92.

Carnap, Rudolph 1956. *Meaning and Necessity*. Chicago: University of Chicago.

Church, Alonzo 1950. "On Carnap's Analysis of Statements of Assertion and Belief". *Analysis* 10: 97-99.

Davidson, Donald 1968. "On Saying That". *Synthese* 19: 130-46.

Field, Hartry 1973. "Theory Change and the Indeterminacy of Reference". *Journal of*

[17] Thanks to Jennifer Carr, Paolo Santorio, Stephen Schiffer, Scott Surgeon and Robbie Williams for helpful comments on an earlier draft.

Philosophy 70: 462-81.

Field, Hartry 2001. *Truth and the Absence of Fact*. Oxford: Oxford University Press.

Field, Hartry 2009. "Pluralism in Logic". *Review of Symbolic Logic* 2: 342-59.

Gordon, Robert (1986). "Folk Psychology As Simulation". *Mind and Language* 1: 158-71.

Grover, Dorothy, and Joseph Camp and Nuel Belnap 1975. "A Prosentential Theory of Truth". *Philosophical Studies* 27: 73-125.

Lance, Mark and John O'Leary-Hawthorne 1997. *The Grammar of Meaning*. Cambridge, UK: Cambridge University Press.

McGee, Vann 1985. "A Counterexample to Modus Ponens". *Journal of Philosophy* 82: 162-71.

Quine, Willard 1953. "The Problem of Meaning in Linguistics". In his *From a Logical Point of View*. Cambridge: Harvard University Press.

Quine, Willard 1960. *Word and Object*. Cambridge, MA: MIT Press.

Quine, Willard 1970. *Philosophy of Logic*. Englewood-Cliffs: Prentice-Hall.

Schiffer, Stephen 1987. *Remnants of Meaning*. Cambridge, MA: Bradford.

Schiffer, Stephen 2004. *The Things We Mean*. Oxford: Oxford University Press.

Shapiro, Stewart 2003. "The Guru, The Logician, and the Deflationist". *Nous* 37: 113-32.

Speaks, Jeff 2014. "What's Wrong with Semantic Theories Which Make no Use of Propositions?" Chapter 2 of Jeffrey King, Scott Soames and Jeff Speaks, *New Thinking About Propositions*. Oxford: Oxford University Press.

Tarski, Alfred 1956. "The Concept of Truth in Formalized Languages". In his *Logic, Semantics and Metamathematics*. Oxford: Clarendon Press, 152-278.

Williamson, Timothy 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

Wilson, Mark 1982. "Predicate Meets Property". *The Philosophical Review* 91: 549-89.