Saying without Knowing What or How

Elmar Unnsteinsson

Postprint. Please cite published version:

Croatian Journal of Philosophy (2017) 17(3): 351–382

Link to published version

Abstract

In response to Stephen Neale (2016), I argue that aphonic expressions, such as PRO, are intentionally uttered by normal speakers of natural language, either by acts of omitting to say something explicitly, or by acts of giving phonetic realization to aphonics. I argue, also, that Gricean intention-based semantics should seek divorce from Cartesian assumptions of transparent access to propositional attitudes and, consequently, that Stephen Schiffer's so-called meaning-intention problem is not powerful enough to banish alleged cases of over-intellectualization in contemporary philosophy of language and mind.

1 Introduction

Many linguists and philosophers of language believe there are linguistic expressions which are phonetically unrealized. Such expressions are syntactically real but lacking in phonetic and phonological properties. One of the most theoretically entrenched examples is (big) PRO which, according to current linguistics, occurs silently in sentences like

(1) $[_{S}[_{NP}$ Wanda$^{1}][_{VP}$ wants PRO$_{1}$ to win$]]$

and is anaphoric on its head NP. Clearly, the postulation of a silent expression like PRO raises all sorts of fascinating questions, some of which have been of particular interest to philosophers.

*(✉) elmar.geir@gmail.com

*Many thanks to Stephen Neale, Thomas Hodgson, Daniel Harris, Guy Longworth, Jessica Keiser, Una Stojnic, and Michael Devitt for helpful comments and suggestions about this paper. I am also grateful for the support of the Icelandic Centre for Research (163132-051) as well as the Irish Research Council (GOIPD/2016/186) while researching different parts of this article.

Stephen Neale, in 'Silent Reference' (2016), does an excellent job of bringing the questions and issues involved to the fore.[1] He is particularly concerned, as I see it, with showing that philosophers ought to be more careful and discerning in their use of this instrument in theorizing. Surely, his advice should be taken to heart. Philosophers need to consider when it is appropriate and plausible to posit phonetically unrealized expressions or syntax and when it is not. It may, for example, be all too tempting for, say, an epistemologist to say that speakers simply refer implicitly to epistemic standards whenever they use the word 'know.' But this raises all sorts of questions. How do they do so? Are they aware of doing it? And are they aware of doing it in the way the theory says they are?

However, in this paper, I argue that Neale's basic metaphysics is too restrictive to do justice to the theoretical options open to philosophers. He assumes, specifically, that it would be absurd to entertain the possibility of *uttering* phonetically null expressions. He also defines the class of aphonics of interest as expressions which, 'by their very nature,' lack phonological properties. I argue that these are mistakes and, further, that they are inconsistent with Neale's other commitments. And those other commitments are, by the look of it, more important. In the final section, I argue that Stephen Schiffer's so-called meaning-intention problem and Neale's related aphonic-intention problem are considerably less serious than they suggest. Borrowing a page or two from Peter Carruthers' (2011) work and from research in dual system psychology, I show that Schiffer and Neale make doubtful and controversial assumptions about the reliability or transparency of speakers' self-knowledge, making the meaning-intention problem far less effective in combating the alleged over-intellectualization of other theorists. Importantly, however, I argue that Gricean intention-based semantics can easily survive as a Cartesian divorcee, since meaning can still be determined by speakers' communicative intentions; they just don't necessarily have conscious awareness of the contents of those intentions.

2 Mad Hatters, Cheshire Catters, and Troublemakers

According to Neale, there is implicit reference and indirect reference. An object is referred to indirectly when a proposition which is merely implicated by a speaker has an object dependent truth condition. Implicit reference, however, occurs when a speaker expresses an object dependent proposition without there being any particular linguistic expression with which reference to the object is achieved. So, for example, if

[1] Page numbers in parentheses refer to Neale's paper unless indicated otherwise.

some philosophers are to be believed, and speakers can intend to refer to the location of the rain by merely uttering

(2) It's raining

on a given occasion, then implicit reference is indeed ubiquitous in linguistic communication. In very general terms, there are two schools of thought on the nature of implicit reference: The Mad Hatters and The Cheshire Catters. The Hatters (for short) believe speakers can refer without there being anything at all with which they refer. They are as mad as a hatter, speak in riddles, expect their audience simply to work out what they intend and, just like the Mad Hatter, are punished *before* committing a crime rather than after (it's all in the pragmatics, you see). The Catters don't speak in riddles but they see non-existent objects everywhere, such as smiles and aphonic variables. In particular, they pretend to see these objects even when there is no theoretical need to do so.

Now, more precisely, the Catters are philosophers who wish to posit aphonic syntactic material in order to explain any plausible case of implicit reference. So, for example, just like linguists want to introduce the aphonic PRO in (1), Catters might propose to introduce an aphonic location variable in (2), which could give us (3)

(3) $[_{S}[_{NP}$ It$][_{VP}$'s raining$[_{PP/AdvP}\ x]]$

as a possible syntactic representation of (2).[2] On this model the aphonic variable could be occurring as an NP within a larger PP (substitutable for 'in Dublin') or as an AdvP (substitutable for 'here'). In this case, introduction of the variable is motivated, most obviously, by the claim that a location is necessary for an utterance of (2) to be evaluable for truth or falsity and, also, by the idea that the variable could be bound by an explicit quantifier, as in 'Everywhere I go, it's raining.' Mad Hatters like Neale and Schiffer, however, consider it much more important that ordinary speakers actually see themselves as having intentions to refer to a location when uttering a sentence like (2). More about that particular madness later (§4).

It seems like Neale wants in some sense to be both a Hatter and a Catter, so he takes on the role of Alice in 'Silent Reference,' trying to make sense of all the strange things in Wonderland. He tries to make the debate between Hatters and Catters more precise and starts by pointing out certain limitations of being a Catter. He points out

[2]Note, however, that almost everything about (3) is controversial because, for one, expletive 'it' is here either a non-argument or quasi-argument. If it is construed as a non-argument – as in constructions like 'It seems that ...' – the gerundive 'raining' in (2) ought to be analyzed as CP with empty complementizer. It's also worth noting that many theorists would propose much more complicated analyses of a sentence like (2), involving multiple hidden variables—for time of utterance, the utterer, the world, etc.—I focus on the location variable here for simplification (see, e.g., Lewis 1970).

that aphonic expressions like *x* in (3) and (unwritten) in (2), will have some rather strange features. First, they are proper parts of the sentences in which they occur but they never correspond to any part of any utterance of the sentence. This makes them very different from expressions like 'cake' and 'eejit.' Secondly, he argues, on this basis, that there can be no such thing as compositional semantics in which the meanings of parts of utterances compose to yield meanings for whole utterances, if one of the parts is supposed to be an utterance of an aphonic expression. A whole utterance of a sentence is a sequenced event which can be segmented into sub-events where each sub-event corresponds roughly to a word in the sentence. And, again, if 'eejit' is part of the sentence uttered, there will normally a be a roughly demarcated part of the utterance-event which corresponds to that word. No such utterance-parts will be found to correspond to PRO or *x* in (1) and (3). Therefore, compositional semantics cannot take as inputs the meanings of parts of *utterances*, if the semantic properties of aphonics are to play any role in composition. Composition must take as inputs the semantic properties of something other than utterances of expressions, it would seem; perhaps the expression-types themselves.

It would seem to follow, then, that one can't like utterance-based compositional semantics while being a Catter. But, of course, there are those who appear to do exactly that and we can call them Mad Catters (Stanley 2007 and Recanati 2010 are possible examples). After looking around for truthmakers to make Neale's two claims true, I realized I could find nothing but troublemakers. In what follows I discuss two such troublemakers before, in the next section, turning to more specific arguments against Neale's position. We should all be free to be Mad Catters when I'm done.

2.1 Omissions

Neale is rightly concerned with spelling out the nature of and connections between *words*, *sentences*, *utterances*, *propositions* and so on. Words are abstract artifacts created by linguistic communicative acts and sentences are then, presumably, abstract structures suitable to contain such artifacts in various syntactic arrangements. On Neale's view, utterances of sentences are events. Specifically, they are events whereby sentences are represented or, as he likes to put it, utterances are *proxies* for sentences. He makes the important point that the traditional distinction between expression-types and expression-tokens blurs and confounds these more fine-grained distinctions. There are not two fundamentally distinct kinds of linguistic expressions, i.e. types and tokens; there are, rather, expressions and various kinds of proxies for those expressions. A somewhat similar point has been made before (Searle 1978; Kaplan 1990) but the distinction still looms large in the literature and Neale makes particularly clear how detrimental to good theoretical sense it can be. Crudely put, utterances or

inscriptions of sentences are not sentences any more than a picture of the Queen is the Queen.

Neale's discussion of aphonics would have been helped, though, by a more detailed examination of the kind of event an utterance or inscription is. As he is most certainly aware, utterances (let's ignore inscriptions for now) are events under intentional description as their source is an intentional agent with goals, reasons, desires, beliefs and various cognitive and circumstantial limitations. In brief, they are intentional actions. Relatedly, the interpretation of an utterance by a normal hearer is geared towards the event as an intentional action: why did they choose those words? why are they saying what it seems like they're saying? Interpretation is geared towards reason-based explanations of intentional action. Linguistic interpretation—interpretation of speech acts—is just a special case of attempts at action understanding more generally. We automatically and effortlessly interpret human actions in terms of beliefs, desires, and intentions (e.g. Carston 2002: 42-44). Seeing someone walking repeatedly over some area in a field, their eyes moving quickly from one part of the grass below to another, I immediately assume they *want* to find something they lost, and that they *believe* it is there somewhere.

Already, this is a prima facie troublemaker for Neale's argument against Mad Catters. For ease of exposition, let's use 'action' for a complex intentional action and 'act' for any proper part of such a complex. What I mean by 'proper' part here is that the part is intentional just as much as the more comprehensive action of which it is a mere part. So, when I intentionally bake a cake, the act of breaking the eggs is an intentional proper part of the more comprehensive action. According to some philosophers, there are actions and acts that have no spatiotemporal properties at all. These are so-called acts of omission or refraining. Randolph Clarke (2010, 2012a, 2012b, 2014) argues, for example, that some omissions consist in the total absence of action relative to an agent, time, and location. Omissions can be unintentional or intentional; the latter he calls refraining. One further condition on refraining to V, for Clarke, is that there is some norm, standard, or ideal in place to the effect that one should V (2014: 29). There are others, however, who argue that refraining is always a type of action (e.g. Brand 1971; Fischer & Ravizza 1998). So, if an MP chooses to refrain from voting on a bill in parliament some particular bodily movement – or, even, the act of keeping still exactly then and there – must constitute the act of refraining at that time and place.

In fact, it doesn't matter what ontology of omitting or refraining we commit to, Neale's argument can only be saved if he can show that Mad Catters are, for some reason, not allowed to appeal to these notions in welding together aphonics and utterance-based compositional semantics. All parties to the debate agree that there is a sense in which people can intentionally omit to do something. Kent Bach (2010: 54-56) insists

that, still, there is no sense in which refraining or omitting can count as actions or acts. But, as he realizes, omitting is not simply *not doing*. What counts as an omission, Bach agrees with Clarke, "is itself partly a normative matter" (Bach 2010: 54). So, whatever else it is, refraining from acting is part of folk psychology on all fours with acting, speaking, expressing, and so on. That is to say, even if refraining to act is not really to act at all or consists in the absence of particular acts or actions, these non-acts can figure significantly in speech acts, their planning, and in the interpretation of speech acts. For our purposes, then, there is no harm in calling omission and refraining acts or actions. Just bear in mind that they could turn out to consist in the absence of an action or act on a given occasion – and so, strictly speaking, they are not actions or acts – or, alternatively, they correspond to something that was actually done. All we would need to do to accommodate Bach's insistence is to say that understanding intentional behavior in general is directed towards two kinds of objects; action and inaction.

To be clear, I am arguing that refraining to act on an occasion is in perfectly good standing, on anyone's account, when it comes to the automatic attribution of mental states to intentional agents in explaining their behavior. When I see someone accidentally drop a penny while walking in high-grown grass, stopping only for a fraction of a second to gaze down, I immediately assume they believe they lost a penny and that it is pointless to look for it. Arguably, Pennyless (their name) refrained from searching and I, watching, automatically explained this fact to myself by assuming various things about their mental state. If asked, Pennyless might confirm that searching for the coin would have been pointless, hence better to decide to do nothing at all. If I were Pennyless I would have done the same, I might think, and doing the same is the doing of nothing.

If refraining figures in action explanation generally, it also figures in utterance explanation in particular. As Clarke (2014 : 32) points out, one of Strunk and White's famous dicta in *The Elements of Style* was "omit needless words." Clarke adds that, whenever one complies with this stylistic norm, one brings about the omission of words by the act of omitting their use. Furthermore, syntactic structure itself provides for a wealth of low-grade normative properties to capture the sense in which speakers refrain from uttering one thing in uttering another. So, for example, when I utter (2) while in Dublin it's clear to all that I should have added 'here' or 'in Dublin' if I wished to be more explicit, and if indeed my plan was to talk about the weather where I was located. There is a longer construction which I should have used in case I believed the context called for it. Let's say, then, that I refrain from saying explicitly where it rains in uttering (2). My refraining either consists in the absence of an act or it consists in some short-lived or instantaneous movement or other; quick breath, glance, gesture, whatever.

We have, then, candidate acts for being parts of utterances corresponding to aphonic parts of sentences. Utterances are actions which can, on occasion, be partly constituted by acts of refraining from saying something explicitly. Moreover, speakers can easily report on their acts of refraining after the fact. MPs may abstain from voting and report this by raising a hand or saying "I abstain". On some views, these latter actions would actually be spatiotemporally constitutive of the act of abstention but, as I have said, my argument doesn't require this assumption. I conclude that it makes perfectly good sense to say that, on occasion, speakers will intentionally perform the act of omitting to say explicitly. They do so, for instance, when they utter (2). This suffices to make trouble for Neale's argument. We have found a candidate to be the utterance-part corresponding to any plausible aphonic sentence-part. The candidate plays a significant role in speakers' capacities for mindreading, communicating, interpreting and explaining intentional action more generally. We could even imagine the communicative defects one would incur if one were, so to speak, omission-blind, and could only ever understand action, never inaction. Surely, this would be debilitating. And, finally, if there is an aphonic location-variable in sentence (2) it can correspond to the act of refraining from referring explicitly to a place.

2.2 Gaps

And what's the problem with instantaneous or durationless proper parts of utterances anyway? As research in phonetics and phonology shows, the correspondence relation between the abstract sentence or word and their audible utterance proxies is extremely complex and counterintuitive. This work has, for example, revealed what is sometimes called the 'lack of invariance problem,' namely that there is no one-to-one correspondence between acoustic signals and perceptual categorization into phonetic segments. In speech perception, different acoustic patterns are invariably perceived by hearers as a single pattern. Speech perception is 'categorical' on this way of thinking. The main reason for this is the phenomenon of coarticulation: the fact that discrete segments of speech are influenced acoustically by the immediately preceding or following sounds uttered. Take the articulation of the /p/ segments in 'pole' and 'peel.' The different positions of the lips, which is explained by the difference in the following vowels, creates differences in the acoustics. Normally, however, this difference is not reflected in the hearer's perception of the utterance (Unnsteinsson 2017).

Another counterintuitive feature, more relevant to our concerns, is that coarticulation occurs both within words and across words in flowing speech, resulting in the fact that, most of the time, gaps between words are not indicated by the continuous speech signal at all. Obviously, this is where inscription is usually very different. So, to take Neale's example, in uttering (4) the speaker sequentially produces five

word-occurrences although the sentence contains only four words.

(4) The cat ate the mouse

Now, let's just assume the phoneticians and phonologists are right about all of this. This creates well-known problems about how knowledge of word boundaries is acquired so quickly and effortlessly by children. Yet such knowledge can elude adults for a long time as well, for particular expressions (looking online for two seconds I found the example of 'housechablis' instead of 'house Chablis'). But this seems to be another troublemaker for Neale. When competent speakers utter (4) they will utter a sentence containing at least four word-gaps indicating that one word has stopped and another has begun. But the gaps in the sentence will almost never correspond to any identifiable gaps in any utterance of the sentence. The question is: how do speaker/hearers then learn where one word ends and another one begins? Presumably the answer has something to do with learning to identify similar acoustic patterns in different linguistic contexts. For example, competent speakers will also understand an utterance of, say, 'The mouse ate the cat.' But, of course, it will still be the case that almost no particular utterance of (4) contains parts corresponding to the word-gaps. So, competent speakers can utter a sentence with four gaps, understand effortlessly that the sentence has four gaps, without there being recognizable gaps in the utterance itself.

But the trouble doesn't start properly brewing until we reach the level of the phrase marker. If linguists are to be believed, sentence processing in ordinary speaker/hearers must involve the mental construction of an abstract syntactic structure. A fully developed human parser – the internal mechanism for processing sentences – at least assigns structures encoding various dependency relations between words and phrases to sentences encountered in speech and writing. In a recent book, David Pereplyotchik provides a wealth of arguments for the psychological reality of what he calls 'mental phrase markers.' The most telling arguments are based on results from brain-studies in neurolinguistics, so-called structural priming experiments, and on plausible explanations of garden-path effects. The data strongly suggest that there is an independent syntactic processing-stage which occurs before any semantic or pragmatic information is accessed. And this stage involves the construction of mental phrase markers identical to those developed by generative grammarians (Pereplyotchik 2017, Ch. 5).

Take structural priming, for instance (see Pickering & Ferreira 2008 for review). If a speaker encounters and parses a sentence with postulated phrase marker P then, the theory predicts, P has been activated in the speaker's mind. If P is activated it should remain so for a while and should show up in other mental processes. Experiments confirm that there are strong priming effects of this sort. So, if P is primed in sentence perception it becomes much more likely that a P-sentence is produced later,

even if semantically equivalent sentences with other phrase markers are equally or more salient in the context. The human parser, it seems, automatically assigns phrase constituency structure to sentences.

It's important to note that, in spite of this, many theorists would argue either (i) that speakers have no knowledge or beliefs about mental phrase markers; even that they aren't mentally represented, or (ii) that any such knowledge or belief is tacit, subpersonal, subdoxastic, or inaccessible to consciousness.[3] Cognition involving mental phrase markers, on most accounts, does not reach personal-level explanation. This contrasts with speakers' beliefs about what they intend to say or refer to on a given occasion of utterance. For normal humans, it seems trivially true that they know what they mean by uttering something. They seem, at least, to know some substantial part of it, as manifested in the capacity to repeat, clarify, or paraphrase what was meant (although I will criticize this alleged truism in §4). Still, supposing that competent speakers stand in some cognitive relation to mental phrase markers, and that the correct theory of this relation either falls into category (i) or (ii), sentences will have parts with no corresponding parts in the utterance. The most extreme views in category (i) will surely deny the very existence of mental phrase markers, but we can set them aside for the moment (see Collins (2007) for a moderate (i)-type view).

Assume, then, that speaker/hearers have tacit knowledge, at least, of the immediate constituents of a sentence. They automatically process sentences in terms of mental phrase markers. So, speakers tacitly know about NPs and VPs and constituency-boundaries are even posited as parts of the abstract syntactic structure of a given sentence. But what, if anything, is the part of an utterance of a sentence which corresponds to the part of the sentence-cum-phrase-marker that distinguishes the VP and the NP? Given that there are verb-subject-object languages, such as Irish, where the VP is split by the NP in normal word order (Irish is a VSO language), it's unclear how this question could be answered directly. The question falsely presumes that sentence-parts and utterance-parts that go proxy for sentences stand in simple, isomorphic mapping relations. Syntactic theory shows that a lot of material is properly said to be part of the abstract sentence, while having no obvious counterparts in utterances or inscriptions. So, the fact that some words have this feature as well should not be objectionable as such.

[3] See Devitt (2006) for an example of the first kind of view and Dwyer & Pietroski (1996) for an example of the second.

3 Two Arguments for Uttering Aphonics

Neale reports that when he talks of aphonics he is particularly concerned with "… individual expressions that are unpronounced and unheard *by their nature*, expressions that intrinsically lack phonological features or instructions for pronunciation" (236, italics in original). The idea seems to be that positing aphonics wouldn't be theoretically exciting unless the lack of phonetic and phonological properties is essential to the posited expression. Most theorists allow for aphonics in the less substantial sense in which they are actually phonic expressions that happen to be omitted on a given occasion. In VP-ellipsis, for example, it is important that what is elided be identical to the antecedent verb – which can either precede or follow the ellipsis – as in 'Sally drove to England and Joe [drove] to Scotland.' Neale is interested, it seems, in expressions partly individuated by their lack of phonology, so that trying to utter them would always result in the production of some distinct expression.

This is unfortunate for a number of reasons, the most important being the fact that there is no such thing as an intrinsically aphonic expression. Any expression can be uttered, even if it happens to lack phonetic features or instructions for pronunciation. What's more, Neale explicitly recognizes this elsewhere in his paper.

3.1 Uttering

Neale provides a thoroughly intention-based theory of utterance identity. The question at issue is a metaphysical one: In virtue of what facts is a given utterance and acoustic proxy for a given word? Very roughly and relative to "a few reasonable assumptions," Neale writes that "an utterance u produced by *S* on a given occasion is an utterance of expression *e* iff *S* intended *u* to be an utterance of *e*" (265). On the face of it, this view of utterance identity appears to be incompatible with expressions which are aphonic 'by their nature.' We can plug any allegedly aphonic expression into Neale's biconditional and get, as a result, a speaker's utterance of that aphonic expression. So, for example, I can intend to utter the aphonic expression 'PRO' by articulating the sound /pro/ on a given occasion.

What's more, uttering aphonics is in some sense made easier by their lack of phonetic features. There definitely is already a standardized way of uttering the aphonic expression 'PRO,' at least within linguistics. But, for other less entrenched cases, such as location variables, it seems like we can choose any phonemic pattern that would do the job in the context at hand. Or, in lieu of that, one could utter the expression by omission, as discussed above. This fits with a theory of speech errors I have defended elsewhere, which also incorporates a thoroughly intention-based view of utterance identity (Unnsteinsson 2017). Very roughly, the idea is that when a speaker has expres-

sion e_1 as their target but accidentally produces some other expression e_2, the uttered expression will be a misarticulation of the target expression. It was an odd way, and accidentally so at that, of uttering the expression which was the speaker's intended target. Thus, I could intend to utter 'Obama' but accidentally utter 'Osama.' On this theory of speech error, I will have pronounced 'Obama' as 'Osama' on that occasion.

So, it seems like we have two options, neither of which allows for the possibility of intrinsically aphonic expressions. First, we could say that any utterance u where the speaker intends to utter the aphonic e by making u is such that e was in fact successfully uttered. Since there are no instructions for pronunciation one really cannot go wrong and anything one does—provided one has e as one's target in uttering u—will count as an acoustic proxy for the aphonic. Secondly, we could say that the correct pronunciation of an aphonic expression is in fact silence. The instructions for pronunciation indicate that the expression ought to be uttered by making no sound at all. In effect, this is taking lack of phonetic features to be a kind of phonetic feature. But we don't need to choose between the two options, for both support the same conclusion, as noted before. If we take the second option, all utterances of PRO, for example, where the speaker intends to utter PRO by making the sound PRO (or any other sound) will simply count as a misarticulation of the aphonic. Importantly, however, and this is clearly part of the intention-based approach to utterance identity, any such misarticulation will still constitute an utterance of the expression. It follows, then, that intentionalists must believe that aphonics – though they're usually not heard in utterances – can very well be uttered and heard.[4]

But what on earth would it be like for an ordinary speaker to have an aphonic expression as their target? And how could some sound they emit constitute an utterance of that aphonic target? This is part of Neale's aphonic-intention problem, which I will discuss more directly below. Let's consider these questions naively first. It would seem like some attempts by ordinary speakers to make themselves absolutely clear because of prior misunderstanding might be classified as (temporally extended) utterances of an aphonic like PRO. Say I'm planning a road trip with you and Siobhan and we're deciding who shall drive. We've been uttering sentences like 'I want you to drive' and 'Siobhan wants me to drive,' so this syntactic structure is primed and we are thus a bit more likely to misidentify similar structures. Then I say,

(5) You or Jane wants to drive,

trying to transfer the responsibility for driving over to you or Jane. Primed for mis-

[4]For a very different point of view on this, see Hawthorne & Lepore (2011: 460-465) and Lepore & Stone (2015: 217-220).

understanding, you ask: 'We want *who* to drive?'[5] When I respond Jane is no longer present. Somewhere along the way, my utterance-plan goes badly awry, but what comes out of my mouth is something like the following:

(6) *I said you or Jane wants *yours-her-self* to drive

Now, of course it is far from obvious what exactly we should say about this strange case. Perhaps I had the phonic expression 'yourself' first as a target and then, thinking I could fix the error, I had the phonic expression 'herself' as a target. So, didn't I just misarticulate those expressions? Probably, yes. But let's assume PRO exists and is really an aphonic pronoun controlled, in a case like (5), by the subject of the matrix verb, i.e. 'You or Jane.' Add to that the idea that PRO inherits the reference from its antecedent. It's not crazy to suppose, then, that when I made the error and uttered /yours-her-self/ my target was an expression with exactly those features. I just couldn't find any phonic expression in my mental lexicon which corresponded well to those features. The problem is that the subject, 'You or Jane,' can easily control aphonic PRO, but it can become awkward with an overt reflexive pronoun as in (6).

Still, I don't want the argument to rest entirely on the plausibility of this kind of case as it may be judged a bit far-fetched. However, even if ordinary speakers never have an intention to, as we might say, give phonetic realization to an aphonic, it is arguable that experts both could and routinely have such intentions. When linguists or philosophers utter or inscribe 'Wanda wants PRO to win' or 'It's raining in x' one possible description of what they're doing is that they are uttering or inscribing the posited aphonic expression. To support this, I see no logical or metaphysical impossibility in the idea that after a few decades the use of PRO would catch on in the general population. This is, of course, terribly unlikely, and even more so in the cases which are of particular interest to philosophers, like the ideas of aphonic location-variables or modes of presentation.

There is, however, an obvious objection which is implicit in Neale's own paper. As he notes, when linguists write PRO in a sentence, it's part of their structural description of the sentence (243). So, one could say, rather than uttering or inscribing the aphonic expression, what linguists are doing is describing, or perhaps simply naming, the expression. Indeed, since the postulated expression has no phonic properties it stands to reason that the expression is either named or described, not uttered or inscribed, and that this is what the experts intend. I want to fully acknowledge the strength of this point but, it's equally clear, it still doesn't amount to showing that aphonic expressions like PRO or variables for locations are aphonic 'by their nature.' If intentionalism about

[5] If the misunderstanding involved here sounds implausible, just imagine this all happening over walkie-talkie in a movie from the 80s.

utterance identity is assumed, experts can certainly utter these expressions if they want to.

In his discussion, Neale introduces PRO as "silent *self*" which corresponds to what he calls "stilted-'self'" but he makes clear that he thinks they must be different expressions. And it's clear why he thinks this: one is by its nature aphonic and the other is an odd or stilted extension of the phonic word 'self' into a set of unfamiliar syntactic distributions, namely exactly the distribution of PRO or silent *self*. It follows from Neale's assumptions – although, as already noted, it's not compatible with his notion of utterance identity – that silent *self* and stilted-'self' are different. The former is not stilted and the latter is not silent (243). But why suppose that this is a robust criterion for individuating words? Well, we shouldn't suppose so. Before arguing for this claim, and responding properly to the objection from structural descriptions, we need to go through the second argument for the claim that there are no intrinsically aphonic expressions.

3.2 Uttering What?

Let's agree with Neale that words are abstract artifacts, along with things like laws and conventions. Agree also that word-proxies or tokenings of words are not words (see his Section 6). But can the nature of words as abstract artifacts be described in more detail? It would appear so. Wolfram Hinzen and Michelle Sheehan (2015) argue, for example, that there are four important notions of 'word' all of which play significant roles in current linguistics.[6]

> First, there is the notion of the word as a *prosodic* unit (the 'phonological word'); then there is the notion of the semantic word or *lexeme*, which is the word understood as an abstract vocabulary item with a given meaning that can take different forms, such as the verb RUN, which can take the forms *runs*, *ran*, *run*, etc. Even more abstract is the notion of a lexical *root*, which involves a semantic core possibly shared across lexemes of different categories, e.g. the root √RUN as involved in the verb *run* and the homophonous noun *run*, which occurs in the expressions a *run*, *many runs*, *running*, *Mary runs*, etc. Finally, there is the *grammatical* word: the word as a *morphosyntactic* unit or as functioning in a sentence context. (38, italics in original)

Neale agrees with Kaplan (1990, 2011) and others that the first unit on this list is not what we're looking for when asking about the metaphysics of words. Of course,

[6] Thanks to James Miller for alerting me to this passage (Miller, 'The Metaphysics of Words,' unpublished).

there exists such a prosodic unit, but it's clear that there are psychologically real and important distinctions between phonologically and phonetically identical words. More importantly still, prosodic units can change so dramatically over time that their individuation is problematic. Kaplan (2011) and Hawthorne and Lepore (2011) argue for something like a lexeme-based theory of word identity. On their account, then, words are essentially objects with certain syntactic markers or properties as well as semantic ones: they are verbs or nouns for instance. As Hinzen and Sheehan (2015) point out, the category of lexeme is almost identical to what many philosophers call 'concepts.' Then there is also the more abstract notion of lexical root, which is shared by different lexemes. Finally, there are morphosyntactic units, characterized, for example, by the position the unit occupies in a phrase-structure tree or the manner in which it interacts with various affixes.

This provides a wealth of possibilities for how philosophers could define words. The correct metaphysics of words might incorporate any combination of these four properties, and of course, which ones are appropriated may depend on the theoretical purposes at hand. Most theorists seem agreed that, when individuating words for the purposes of describing the items stored in the mental lexicon, prosodic units are not terribly important. There certainly are prosodic units but they are non-basic. It's reasonable to suppose, then, that a metaphysical theory of word identity will always incorporate at least one of three properties: (i) being lexemic, (ii) having a lexical root, and (iii) having a morphosyntactic profile. Words don't need phonetic realization – because we are assuming that there are aphonics – and whatever phonological properties they have can change dramatically, fluctuate or even disappear. Semantic properties of words allow for similarly dramatic fluctuations over time.

But here Neale's point about the difference between silent *self* and stilted-'self' may seem relevant: Wouldn't aphonic and phonic units always constitute distinct expressions? No, they would merely be distinct prosodic units, because all of the other properties could still remain intact. Again, as a comparison, does an expression with semantic properties become a different expression if it evolves into an expletive, 'non-semantic' unit? For instance, is expletive 'it' distinct from 'it' occurring as argument? The answer to these two questions would depend on whether or not one endorses the lexeme-based view of word identity. But the answer to the former, it seems, has to be a direct 'No.' Phonetic and phonological properties are superficial, non-basic features of words. Therefore, there is no such thing as an intrinsically aphonic expression.

Now we are also in a position to respond to the objection mentioned near the end of Section 3.1 above. According to the objection, when experts appear to be uttering aphonic expressions like PRO, silent-*self*, or variables for location, what they are really doing is providing structural descriptions or merely naming the aphonic item. Well, if both of my arguments are sound, this is seen to be unduly ad hoc. If phonological

properties are superficial and utterance identity is determined by speakers' intentions, nothing hinders experts in intending the production of a particular sound as the utterance of some postulated aphonic expression.[7] Before, I gave reasons to think that if such expressions exist at all, the omission of any phonic counterpart in an utterance may count as an act of uttering an aphonic. I believe it is difficult to find credible cases of non-expert speakers in fact having intentions to utter an aphonic *by producing some sound or other*. Doing so, however, is open to the experts themselves. All I needed to show is that no word – apart from mere 'phonological words' of course – is aphonic by its very nature and, so, it is fine if it turns out that experts normally consider themselves only to be describing or naming aphonics, rather than actually giving voice to them. But when they in fact intend to utter the aphonic by producing some sound or other, the identity of the word in question isn't suddenly altered.

4 What About the Meaning-Intention Problem?

According to Schiffer and Neale, there is a significant difference between theories on which speakers have intentions to refer implicitly to something like a rain-location and ones on which they have intentions to refer implicitly to modes of presentation or epistemic standards. When it comes to rain, speakers seem immediately and effortlessly aware that they in fact intended to refer to a location even if they omitted any expression whose function is specifically to enable such reference. Ordinarily, when speakers say that it's raining, they'll have no trouble answering the question 'Where is it raining?' if it is somehow unclear in the context. So, it may seem, we have good reason to suppose that speakers actually have intentions to implicitly refer.

Modes of presentation appear to be different. Many philosophers wish to posit modes of presentation to solve puzzles about singular reference. According to one influential theory of this sort, when speakers attribute beliefs to others with sentences like (7),

(7) Bianca believes that the hippopotamus is sleeping,

they really express the proposition that Bianca believes, under some mode of presentation, that the hippopotamus is sleeping. So, uttering a sentence like (7) involves implicit reference to something called a mode of presentation (see Neale 2016, §14.4 for details). But, on the face of it, speakers are not aware of intending anything of

[7] There is one other possibility, however. One might simply argue that aphonics aren't expressions at all, that they are much more like phrase structure trees, Case, and other parts of the syntactic description of sentences. I find this possibility appealing, but won't address it here, as it would take us too far into different territory.

the sort. If the speaker were asked to specify which mode of presentation they had in mind in uttering (7) the question would usually not be understood. Maybe there are more intuitive ways to get at the question, which would better track what theorists have in mind by positing these modes, but still, the difference between this case and the first one will remain. Normal speakers are completely aware that they implicitly refer to locations all the time, but, if they do indeed refer to modes of presentation – or epistemic standards, according to Schiffer – they haven't the faintest idea that that's what they're doing. It follows, then, that only cases of the first kind are compatible with the assumption that speakers have transparent, privileged access to what they consciously mean, intend, and believe in uttering something. And it is reasonable, Schiffer contends, to think that such access is "part of a normal person's functional architecture" (1992: 515; Neale 2016: 320).

Neale argues that aphonic reference presents an even deeper problem (2016, §14.7). Indeed, if there are theorists who hold that speakers refer to modes of presentation *with* aphonic expressions, it follows that they refer to things they don't know about with things they don't know about. So, Neale asks, is it really plausible to attribute aphonic-involving referential intentions to speakers at all? Keeping with the example from before, it is fairly clear that ordinary speakers don't appear to have conscious beliefs or intentions about things like PRO. Linguists didn't discover PRO until very late in the history of natural languages where PRO is allegedly part of some sentences, so speakers' access to expressions like PRO is very different from their access to expressions like 'eejit.' The logical conclusion, then, seems to be that positing aphonic-involving intentions is also incompatible with Schiffer's assumption that normal speakers have privileged access to what they mean, intend, and believe in uttering something.

Obviously, then, the degree to which these problems are worrying should match the degree to which the transparency assumption is in fact justified. More precisely, there are at least three hidden assumptions at work here:

> A1. Speakers have transparent, privileged access to what they consciously intend, mean, and believe in uttering something.
>
> A2. What speakers consciously intend, mean, and believe in uttering something is identical to what they really intend, mean, and believe in uttering something.
>
> A3. Interpretation and inference are not necessary in understanding what one oneself intends, means, and believes in uttering something, but they are necessary in understanding what others intend, mean, and believe in uttering something.

Intention-based semantics, as promulgated by Neale and Schiffer, is thoroughly wed-

ded to A1-A3. This is unfortunate, not only because it will appear to critics that there is no such thing as intentionalism without broadly Cartesian views of the mental, but also because there are good reasons to commit transparency to the flames. Or so I will argue, along lines essentially similar to Neale's "Tacit States Reply" to the meaning-intention problem (2016, §14.12).

4.1 Aphonic-Involving Intentions

Let's start with the alleged problem with aphonic-involving intentions before approaching the broader question of meaning-intentions. As stated, the problem is that speakers would be supposed to perform acts of meaning, saying, referring, etc., without knowing anything at all about the means with which they do so. Clearly, PRO and other aphonic expressions will fit the bill; normal speakers don't appear to have any conscious knowledge that such entities exist. On the other hand, speakers appear to know that words exist and they appear to know that words and sentences are means by which they express and communicate their thoughts and beliefs to others.

That may indeed sound reasonable, but only because it is part of folk linguistics and general common sense opinion. If any account of the nature of words along the lines of §3.2 above is correct, normal speakers have no idea what a word is, because they have no idea what a lexeme, root, or morphosyntactic unit is. But this objection is too quick. Surely, no one would suggest that naïve speakers know the true metaphysics of words, what they do know is that there is something in the world, namely utterances of words and sentences, with which people communicate. And nothing similar can be said about their knowledge of utterances of aphonic expressions.

The deeper problem with this argument is that it draws an illicit distinction between words on the one hand and aphonics on the other. For if we are to accept aphonics at all, they will belong in the category of words; they are merely words without phonetic or phonological properties. So, if we really want to take folk linguistics seriously, naïve speakers will turn out to know about aphonics in virtue of their admittedly superficial knowledge of words in general. Neale, it seems, would have to draw some principled distinction between phonic expressions and aphonic expressions. This would allow him to hold that normal speakers have only encountered utterances of the phonic bit of the lexicon and, so, they only have knowledge of words as phonic entities. Aphonics are completely beyond their ken.

Apart from the problems already mentioned, especially the point that normal speakers may have encountered aphonics in speech by witnessing acts of omitting to say something explicitly, drawing this distinction is not as easy as it seems. To see this, consider some of the most basic theoretical commitments of linguists who pursue an intentionalist theory of phonological competence. Bromberger and Halle (2000),

for example, argue that phonological descriptions or derivations of dated utterances should not be understood merely as phonetic transcriptions of a speech event—i.e. symbols encoding articulatory movements—but as standing for a sequence of intentions which give rise to those movements. Simplifying dramatically, the IPA transcription [ɹ̟ʷɛd], occurring in a phonological derivation of an utterance, doesn't merely record the phonetic segmentation of an event of uttering 'red' into the three stages of (i) labialized postalveolar approximant, (ii) open-mid front unrounded vowel and, finally, (iii) voiced alveolar stop. It represents the phonetic intention that called for this complex sequence of articulatory movements and positionings. This kind of intention is grounded in linguistic competence, since speakers can very well make the requisite sounds and movements encoded by [ɹ̟ʷɛd] without having the intention to utter a word of English, for example if they don't know the language but just happened to produce the sound in the manner required. Phonetic intentions are intentions, then, to produce speech sounds of specific languages.

But what, more specifically, are the objects of phonetic intentions? According to Bromberger and Halle (2000: 26-27), speakers must, at least, have intentions to produce morphemes. A morpheme is either a stem or an affix of a word; 'an-arch-ic', for example, has two affixes, 'an' (prefix), and 'ic' (suffix), and one stem, 'arch'. A single stem, on this kind of theory, can be pronounced differently in different linguistic environments. The stem 'sell' is sometimes pronounced /sold/ and sometimes /sells/, depending on tense and Case agreement. So, whenever I intend to produce a phonic word I retrieve information about each morpheme from memory, and utter the resulting combination or transformation. Now, we have already attributed intentions to speakers which will be completely unrecognizable to them. Normal speakers usually do not, and need not, know anything at all about the morphemic structure of the words they use. Neither do they need to know that morphemes exist; and they normally don't know. Nevertheless, very basic commitments in (some of) phonological theory involve the attribution of intentions to pronounce morphemes to ordinary speakers. It seems, then, that the assumption of transparency would require a wholesale rejection of these ideas, or very radical revision of foundational assumptions. Neither option is appealing, since dropping transparency seems the easier thing to do (see next section).

Before moving on, it should be noted that adding just one layer of complexity into this analysis will make phonetic intentions to utter phonics appear just as problematic as phonetic intentions to utter aphonics. Plural and past-tense affixes in English have radically different phonological features in different environments. Consider the examples from Bromberger and Halle (2000: 27):

> Plural morpheme: cat/s, child/ren, kibbutz/im, alumn/i, stigma/ta, geese, moose

Past-tense morpheme: bake/d, playe/d, dream/t, sol/d, sang, hit

To explain these irregular affixes, they propose a category of abstract morphemes symbolized with the letter Q. Q has no direct phonetic interpretation but, as I understand the idea, it encodes information about how the morpheme is pronounced (it is an 'identifying index'). Bromberger and Halle take care to note that 'Q' is part of the notation of the theory, and not a symbol that really occurs 'in the mind' of the speaker. But this is at best an admission that we simply lack knowledge in this area, since they expect more work in linguistics will eventually reveal mental structures corresponding to representations in the notation of the theory (2000: 26n10). The bottom line, however, is that speakers of English do utter plural and past-tense morphemes, and they must then, in some sense to be explained, intend to utter Q. But it appears that Q either doesn't encode any phonological information, because the phonetics of Q are so radically dissimilar on different occasions of utterance, or the information is almost impossible to specify, even theoretically. Further, it is part of one fairly influential theory of phonological competence that when speakers intend to utter some phonics, they must intend to utter abstract morphemes like Q. Thus, I conclude, it has not been shown that intentions to utter aphonics have to be more problematic than intentions to utter phonics.

4.2 Access to Propositional Attitudes

The argument, so far, may seem to amount to no more than simple buck-passing. Surely, one would like to say, there is a world of difference between conscious, personal-level intentions like the intention to mean that *p* by uttering X and any subpersonal intention or other mental state involved in knowing the grammar of a language. And so, the thought continues, we haven't solved any problem by attributing, say, phonetic intentions to speakers and hearers. Indeed, there is much reason to think that this is merely a *façon de parler* awaiting elimination when science progresses (Collins 2007, 2008). True, this is a popular and plausible way of thinking about these issues but, I want to argue, the problems involved in personal-level intention attribution are much more pressing and consequential than many philosophers of language have hitherto allowed for. Possibly, speakers are built to consciously represent themselves, to themselves, as having certain intentions and beliefs, without this being a good indicator that those are the intentions or beliefs that they actually have. If so, talking in terms of 'conscious intentions' is also just a manner of speaking awaiting elimination.

Start with the nutshell description of Gricean intention-based semantics, or intentionalism for short, coming from Neale, Schiffer and others. According to intentionalism, what the speaker says and means by making an utterance on an occasion

is metaphysically determined by certain specific audience-directed intentions. What is strictly *said* by the speaker may need to conform, in some sense, to the linguistic meaning of the sentence uttered, but what is otherwise meant – e.g. conversational implicatures – can roam freely from any such constraint. So, for instance, if I utter,

(8) Meet me at the bank,

the only facts that can determine whether I meant a river bank or a financial institution are facts about my communicative intention at the time of utterance. The context could be such as to make it appear to my audience that I meant the river bank – if I'm holding a fishing rod, for example – even if I really intend to refer to a financial institution, and so they might very well misunderstand me. And in many such cases the responsibility for the misunderstanding falls squarely on the speaker's shoulders; they failed to take the full context into consideration before speaking. But it is still the intention that determines which interpretation is the correct one. Neale (2005: 179-180) describes this in terms of an epistemic asymmetry between the speaker and hearer. Speakers *know* what they mean and the hearer's job is to *work it out*. The speaker normally doesn't need to work this out, they simply know what they mean without interpretation or inference.

Already, I believe, it is important to pry this apart. Even if it is conceded that speakers normally know what they mean, or some part of it, this must be flagged as a thesis in the epistemology of interpretation. According to intentionalism, what I said in uttering (8) on an occasion is determined by my communicative intention, it is not determined—except in the epistemological sense of that word—by what I believe I intended to say. Neither is it determined by what I believe my intention is while uttering (8). Sure, I could find out what my intention was in uttering (8) by forming a belief about the intention. The difference between the hearer and myself here is, at least, that I often have access to more data – my own mental imagery at the time for example – and may often form my belief without waiting to hear the words I utter. Perhaps this higher-order belief is formed automatically and unconsciously, but it can surely be mistaken like any other belief. Note, however, that more data does not necessarily result in more reliable judgment, as it might just overwhelm one's cognitive system and lead to an increased number of errors. Sometimes, cognitive processes are more reliable if they use only a limited collection of evidence (Gigerenzer et al. 1999; Carruthers 2011: 24). Further, as Peter Carruthers (2011, Ch. 2.5; also pp. 12, 22, 70) has argued, Cartesian transparency derives no support from phenomenological observation, contrary to widespread opinion. Beliefs about what others mean in uttering something are formed just as automatically and sub-consciously as beliefs about what we ourselves mean, intend, or believe. And we have often formed those beliefs as hearers, automatically and predictively, before the speaker puts a stop to the

sentence or even before they say anything. And this is predicted by so-called 'forward models' of human cognition (e.g., Pickering & Clark 2014). Normally, people simply find themselves with beliefs about what speakers meant, with no insight into how exactly the belief was formed. We seem also simply to find ourselves with such beliefs about our own propositional attitudes.

4.2.1 Arguments from Modularity

As Carruthers (2011) argues at length, there are good empirical and theoretical reasons to think that the human mind—or a specific mindreading module in the mind—automatically adheres to cognitive procedures which assume the mind is transparent to itself. If so, it is just a near-universal assumption of humans that if one thinks one is in mental state M then one must actually be in mental state M and, also, that if one thinks one is not in mental state M then one must really not be in that mental state (ibid., p. 12). This cognitive procedure is compatible with the suggestion that, in fact, people routinely confabulate and misinterpret their own mental states. And experimental findings indicate that when such confabulation occurs, people don't have subjective access to the information that their self-attribution of a mental state was a complete fabrication. Carruthers takes, as an example, research on commissurotomy patients, where different stimuli are presented to the two hemispheres at the same time.

> The patient fixated his eyes on a point straight ahead, while two cards were flashed up, one positioned to the left of fixation (which would be available only to the right hemisphere) and on to the right of fixation (which would be available only to the left hemisphere). When the instruction, "Walk!" was flashed to the right brain, the subject got up and began to walk out of the testing van. (The right hemisphere of this subject was capable of some limited understanding of words, but, had no production abilities.) When asked where he was going, he (the left brain, which controlled speech-production as well as housing a mindreading system) replied, "I'm going to get a Coke from the house." This attribution of a current intention to himself was plainly confabulated, since the actual reason for initiating the action was accessible only to the right hemisphere. Yet it was delivered with all of the confidence and seeming introspective obviousness as normal. (2011: 39-40)

Even if patients are reminded, and made fully aware, that the surgery can have effects on their access to their own mental states, they still insist that they know for sure what they really intend to do. As Carruthers emphasizes, this does not support total

skepticism about self-knowledge, but it does show that confabulated mental states will appear just as transparently accessible to us as their authentic counterparts.

Now, let's look again at the transparency assumptions A1-A3, starting with the final one.

> A3. Interpretation and inference are not necessary in understanding what one oneself intends, means, and believes in uttering something, but they are necessary in understanding what others intend, mean, and believe in uttering something.

As already noted, this assumption gets no support from intuition or first-person phenomenology, since the cognitive process by which we form beliefs about the intentions of others is just as immediate and automatic as that by which we form such beliefs about ourselves. But more substantively, as Carruthers (2011) argues, if one likes the idea that the mind houses a mindreading module specifically geared towards the task of attributing mental states to intentional agents, some very serious empirical arguments can be mustered against A3. I will only give the flavor of these arguments here, and go on to focus more directly on assumptions A1 and A2.

First, assume that there is a mental module for mindreading. Secondly, we can then ask: What is our best theory of the nature and evolution of this module? This is where Carruthers would introduce his interpretive sensory-access (ISA) account of mindreading, according to which the module only has sensory access to its domain, and this access is always – with two notable exceptions, namely aspects of perceptual and emotion-like states (2011, Ch. 4, 5) – interpretive rather than transparent. And so, the ISA theory predicts that when one attributes propositional attitudes traditionally so-called – intention, belief, judgment, desire, etc. – one must engage in interpretation in both self-attributions and other-attributions. This prediction derives some support from the observation that a mindreading module is most likely to have evolved in response to strong social pressures on individuals to acquire capacities to predict and explain the complex and varied behavior of other individuals. If this is in fact the most plausible evolutionary story we can tell, the simplest hypothesis, according to Carruthers, "… is that self-knowledge is achieved by turning one's mindreading capacities on oneself" (2011: 65). And this, in turn, would suggest that our access to our own mental states is, in essence, the same as our access to the mental states of others, namely by means of inferences based on various sensory cues. In our own case, the data pool will usually be very different, however, involving many private mental episodes and access to personal memory.[8]

[8] Of course it should be noted that this story is contested, most obviously by simulation theorists like Goldman (2006) who argue that self-directed metacognitive abilities are evolutionarily prior to other-directed mindreading abilities.

4.2.2 Arguments from Dual System Psychology

Whatever one thinks about modules for mindreading, there is a strong case to be made against Cartesian transparency on the basis of dual system theories in psychology, or 'fragmentational' theories of mind more generally. By now it is a fairly standard view, but certainly not universal, in the cognitive sciences that the human mind has a fast, intuitive, automatic, and nonconscious processing component and a more effortful, reflective, and conscious component which is subject to voluntary control (cf. Evans & Frankish 2009).[9] Call the first 'System 1' and the second 'System 2'. There is much controversy about how exactly the two Systems relate to one another and how they ought to be defined. System 1, or parts of it, is evolutionarily ancient and is shared with some non-human animals. System 2 is thought to be more distinctively human or, at least, more developed in humans than in other animals. Either the Systems are distinct capacities with different evolutionary histories and physical realizations or they are different ways of utilizing the same cognitive resources, with System 2 operations partly realized in System 1 processes (Carruthers 2009; 2011: 98-101).

Either way, however, if mental partition of this sort is granted, we can inquire into the characteristics of content-bearing mental states as they occur in System 1 or System 2 processing respectively. Philosophers have recently, for example, been very interested in cases where people sincerely profess to deeply held attitudes – e.g. egalitarianism, anti-racism – while unreflective, 'System 1-based' behavior and cognition seems to manifest diametrically opposed attitudes (Gendler 2008; Schwitzgebel 2010). The possibility of such cases should not be taken as proof that there are two mental Systems or two fundamentally different species of propositional attitude (Gigerenzer & Regier 1996; Mandelbaum 2016), but they are helpful in understanding the status and interaction of different kinds of content-bearing states of mind. To fix ideas, call contentful mental states occurring in System 1 processes 'A-attitudes' or 'A-states' and those occurring in System 2 processes 'B-attitudes' or 'B-states'. A-states are ancient, automatic, and (perhaps) associative while B-states are bookkeepers who, normally lagging behind, strive to be boss over A-states. Tamar Gendler (2008) calls A-state beliefs 'aliefs' and B-state beliefs 'beliefs' but I prefer my own terminology for sake of generality; now we get to talk about A-intentions vs B-intentions, and so on.

Start with beliefs. Suppose an individual S has the B-belief that *not p* and seems

[9] If such a heavy-duty psychological theory, much of which is hotly contested, is not allowed as an assumption, we could make do with a fragmentational or partitive theory of the mind (as in Lewis 1982; Davidson 1982; Egan 2008; Mandelbaum 2016). All we really need is the idea that propositional attitudes can be causally isolated from one another within a single mind, making it possible for a single mind to harbor inconsistent beliefs or intentions at any given point in time. Many theorists find the System 1/System 2 distinction relatively intuitive, so I use it here.

also to know full well that *S* has that very belief. Of course, *S* need not realize that the belief is a B-belief since *S* may not make the A/B-distinction. It's possible, then, that *S* also has the A-belief that *p*, directly contradicting the content of *S*'s B-belief. For example, *S* could be an implicit or unconscious racist, A-believing that other races are inferior, while B-believing that they are not. *S*'s B-belief will consist in things like *S* consciously and intentionally professing to egalitarian and anti-racist attitudes. *S* may even rehearse, in inner speech, the conscious B-belief that other races are not inferior. But, surely, this is compatible with the opposing A-belief being manifested in reasoning, action, and automatic reactions to relevant situations. Finally, there is no longer a clear sense in which *S* knows that *S* believes that other races are not inferior. For, one might think, genuine belief should require at least the presence of A-belief. We might say instead, in this kind of case, that *S* merely knows that *S* professes to believe that *not p* while really believing that *p*. Often, the real belief will turn out to be 'nonconscious,' but this doesn't seem necessary (Frankish 2016; Hunter 2011).

As Keith Frankish (2004) and Carruthers (2011) have argued, using different terminology, B-beliefs are much more like *commitments* than truthful *reports* of contentful mental states. When I say to myself, or to others, that I believe *p*, I commit myself to this belief. More specifically, if I happen to take such commitments seriously, I measure myself according to the standard of believing that *p*: I aim to make my behavior expressive of the belief, experience disappointment when this fails, and exhort myself to stand by my word. Importantly, however, even if my commitment results in everything appearing as if I really believe *p*, commitment is not identical to belief. Unless, perhaps, one is a full-blown instrumentalist or anti-realist about content-bearing mental states. Let's commit to ignoring such views for the time being. To see the point, one just needs to note the relativity and variability of commitment-attitudes across different individuals and across different times for the same individual. Many people routinely commit to things without having any apparent control over relevant behavior or reasoning. So, it seems, the causal profile of commitment-attitudes depends entirely on which other higher-order attitudes are held by the person in question; in the case of belief it may depend on whether they believe they actually have made the commitment, whether they want, generally, to make good on their commitments, and so on.

This immediately contrasts with the causal profile of belief, where the question of whether one really has a belief does not depend on attitudes of this kind. And, Carruthers (2011: 102-107) argues, the same thought applies to decisions, judgments, wonderings, and supposings. An actual decision to do something should "settle the matter" by itself, while saying to oneself that one will do it—committing to it—only does so in concert with the right beliefs and desires. Surely, committing to the truth of *p* may result, in due course, in the A-belief that *p*. And, generally, self-attribution of

the belief that *p* may be self-fulfilling in that it gives rise to various pressures on one to behave *as if* the attribution is true. But committing and believing are still importantly different. Further, one's commitment-attitudes are seriously vulnerable to systematic and immediate confabulation as well as self-directed propaganda and deception, for example in the service of perpetuating a certain self-image (Wilson 2002).[10] I may commit to *p* being true because of wishful thinking, B-believing sincerely that I A-believe that *p* (assuming I have the concept of A-belief), while not really A-believing anything of the sort.

Now, how could this apply to cases like appearing to know what one means in saying something on a particular occasion? Surely this is different from having automatic or non-introspectible attitudes; when one means that *p* by uttering X one consciously intends to mean that p and has fairly reliable knowledge that this is so, right? Remember, however, that the point is not to show that speakers never know what they mean, only that their access to what they mean is interpretive, inferential and prone to error, just like their access to what others mean. And what I have tried to show is that even if speakers know what they B-intend, it doesn't follow that they know what they A-intend. Further, it is quite likely that the contents of A-intentions differ radically and systematically from the contents of B-intentions, even when both are occurrent attitudes a speaker has in making an utterance. So, the idea goes, speakers may A-intend the proposition that *p* while actually B-intending the proposition that *q*. What counts as successful interpretation, on this kind of view, is not completely obvious. But one possibility is that recognizing the A-intention is sufficient by itself, while recognizing the B-intention is not, because it only gives the hearer knowledge of what the speaker consciously believes about the content of the intention. But surely, this is often a good indicator of belief-contents, or some parts of such contents, and will normally be good enough for purposes of everyday communication. But this is all very abstract. Let's consider two kinds of cases where there is, arguably, an actual mismatch between communicative A-intentions and communicative B-intentions.

Hypnosis. As Carruthers (2011: 342-343; citing Wegner 2002) reports, subjects who are given instructions while hypnotized will invent new intentions when asked to explain why the do what they do after waking from the hypnotic trance. It is likely, then, that subjects form intentions and make decisions to act while under hypnosis, and those intentions remain active when they regain consciousness. The intention, say, to

[10] In Kent Bach's (1981) terminology, believing that *p* is different from thinking that *p*. So, if I believe that *p* but want to deceive myself into not believing that *p* I can fill my mind with thoughts to the effect that *p* is false, whenever I have occasion to consider my belief. This does not necessarily change my belief, but it may result in self-deception, that is, my conscious thoughts and imaginings will only ever suggest that I believe that *p* is false, when I really believe it is true. Commitment is more like merely thinking to oneself than really believing.

open the book on the table when you see it, may have been implanted while hypnotized but, when asked why you opened the book, you will automatically form some confabulated belief; that you've always wanted to read Anthony Huxley's *Illustrated History of Gardening*, for instance.

To control for merely pragmatic reasons for subjects to report some reason or other – when they don't really know why they are performing the post-hypnotic action – Carruthers suggests that even ambiguous actions would lead to error prone self-interpretation. So, out of two possible interpretations, the subject would self-ascribe the action which is more plausible in the context, without detecting any mismatch between A/B-intentions. He proposes an experiment where the bodily movements of the subject will be ambiguous between waving goodbye to someone and waving away a bug. But we can also go back to our 'bank'-example above. Suppose we have a hypnotized subject, Beatrix, who is already fairly likely to take money to the bank, and likely to go fishing on the river bank. Suppose, then, that the hypnotist instructs her as follows: "You're holding your daughter's money box to make a deposit into her savings account. When you meet Abigail tell her you're going to the bank, so she can join you there." Let's also assume that this is sufficient to implant, in Beatrix, the A-intention to tell Abigail to come along to the *financial* bank and that this very intention remains active after she emerges from hypnosis. When she is then placed in a situation where the other sense of 'bank' is much likelier to be at play, the current prediction is that she would form a B-intention to tell Abigail that she's going to the *river* bank. Assume, for example, that Abigail and Beatrix are much more likely to go fishing than to go to a financial institution and that, when they meet, they're close to the river bank and Beatrix is holding her fishing rod (as well as her daughter's money box).

A competing description of this case is, surely, that the conscious A-intention to refer to a river bank always supplants the alleged B-intention. But I see this as no more than banging one's Cartesian head against the wall. That is to say, if intuitions of first-person transparency are not allowed to carry weight, both descriptions are at least prima facie plausible, and the issue should be decided by more general theoretical considerations. So, it seems possible that both Beatrix and Abigail misrepresented and misunderstood the former's (A-)intention when she uttered 'bank'.

Implicature and illocution. According to intentionalism, conversational implicature is determined by the speaker's communicative intention on the occasion of utterance. The question of whether or not I conversationally implicate the proposition in (10) by uttering (9) in a particular context is, thus, answered by finding out whether I actually intended to be understood as talking about an open gas station or a closed one (see Grice 1989: 32).

(9) There is a gas station around the corner

(10) that the gas station is open

Implicature breeds plausible deniability. If it turns out that the gas station is closed and, having been reprimanded for misleadingly implying that it was open, I am free to insist that no such thing as (10) was intended. Importantly, however, I am either telling the truth or not. I may have actually intended to implicate (10), succeeded, and then lied about my intention when I realized (10) wasn't true. Alternatively, I may not have intended to implicate that (10) by uttering (9); that's why my later denial can be plausible, for it could, as far as the hearer knows, be true.

But plausible deniability, in turn, breeds credulous self-deception, or so I argue. If it is easy to deny, when challenged, that something was part of one's intention, it is also easy to believe that it actually wasn't, even when this is a form of self-deception. A single propositional attitude only relates to behavior in concert with other attitudes. So, I will only take the umbrella when I believe it's raining if, among other things, I also don't want to get wet. But suppose I don't really believe it's raining and I erroneously self-ascribe the belief that it's raining. As Carruthers (2011: 94) points out, I can still explain why I didn't take my umbrella, while preserving the basic belief – consisting in a near-universally shared cognitive procedure – that my mind is introspectively transparent to itself. I can simply say, to myself, that I really wanted to get wet or that I briefly forgot about the rain while leaving the house. Similarly, if people are put in a situation where it is better – for their self-image or because of social pressure, say – to form the false belief that they didn't intend to imply something or other, they will probably tend to form exactly that belief, immediately and unconsciously.

Consider, by way of example, an argument between a conservative and a liberal about immigration policy. In conversation, Conn has been advocating tight restrictions on the free movement of labor, while Libby wants to abolish national borders. Without articulating explicitly how the statement relates to the more general issue, Conn says,

(11) But the Polish are hard-working.

In this context, Libby may think that Conn is, by uttering (11), implying or presupposing something like: (i) other groups of immigrants are not hard-working, or: (ii) groups of immigrants should only be allowed to enter the country if they have certain character traits, hard-working being one of them, or even: (iii) non-whites are not hard-working (this would require the conversational salience of non-white immigrant groups). And it is not hard to conceive that Conn really did intend to imply one of (i)-(iii) but, equally, in order to preserve an anti-racist self-image, he could deceive himself into consciously thinking that no such thing as (iii) was really intended. Conn could think to himself before uttering (11): I won't be implying anything like (iii)

because I don't believe (iii). But he might still A-believe something like (iii). So, in such a case, plausible deniability becomes a tool for easy self-deception. Unconscious A-intentions and conscious B-intentions come apart.

Consider also the illocutionary force of an utterance. Rae Langton (2009: 33-34) argues that speakers sometimes perform illocutionary acts they don't actually intend to perform. Taking an example from J.L. Austin (1975), she imagines one man saying to another,

(12) Shoot her,

referring to a woman nearby. According to Langton, the speaker (*S*) may have intended the utterance merely as advice while the hearer (*H*) actually takes it as an order. She argues, further, that since illocutionary force ought to be partly defined in terms of conversational uptake, the act of uttering (12) in this context may objectively have the illocutionary force of ordering, rather than advising, regardless of *S*'s intention. This goes against the basic premise of intentionalism, according to which *H* would simply have misunderstood *S*'s actual intention in taking the utterance as an order, if it was really (intended as) mere advice. And so, on this view, the speech act had the illocutionary force of giving advice.

The point here is not to argue against Langton's description but, rather, to show how the distinction between A-attitudes and B-attitudes affords us with a different perspective on a case like this one. Suppose, for instance, that in some sense *S* is aware or ought to be aware that the utterance might be understood as an order in the context. *S* knows, say, that *H* is somewhat deferential and complaisant to others. Still, *S* could convince themself, even just momentarily, that uttering (12) in this context is merely giving advice, not issuing a command. So, it seems, *S* B-intends the utterance of (12) as advice but A-intends it as an order. If this is possible, Langton's description is partly vindicated, even on minimally intentionalist grounds. That is to say, the speech act was really an act of ordering, even if the speaker consciously thought what they were doing was giving a piece of advice. But, on these assumptions, *S* couldn't have A-intended (12) as mere advice because *S* A-believes that (12) is more likely to be taken as an order.

Developing this point in the detail it deserves will have to wait for a different occasion. We have, however, established so far that assumptions A1 and A2 of so-called Cartesian transparency are not nearly as safe as intentionalists tend to believe.

> A1. Speakers have transparent, privileged access to what they consciously intend, mean, and believe in uttering something.
>
> A2. What speakers consciously intend, mean, and believe in uttering something is identical to what they really intend, mean, and believe in uttering

something.

First, it is doubtful that speakers have transparent access to conscious intentions (A1) but, even if this is conceded, it is considerably more doubtful that speakers' conscious intentions are constitutive of their actual intentions. We tell ourselves all sorts of things, and appear to act for all sorts of fabricated reasons, without thereby having transparent knowledge of the contents of our actual propositional attitudes.

Where does this leave us? Well, intention-based semantics, as far as I can see, still stands as a metaphysical theory of what grounds or determines the proposition(s) expressed by a speaker in making an utterance on a given occasion. This will be the speaker's communicative intentions, however exactly those are spelled out in the final theory. Still, since transparency is a doubtful epistemological thesis, the meaning-intention problem becomes less of a sweeping tool for eliminating implicit reference than it appeared to be. Surely, the fact that speakers believe that they implicitly refer to locations in saying things like 'It's raining', still constitutes some evidence that this is indeed part of their communicative intention. And, conversely, their apparent lack of awareness of intending to refer implicitly to epistemic standards or modes of presentation gives some reason to think that they don't. This is simply because people do know something about what they mean, intend, and believe. But they also know something, and in a similar way, about what others mean, intend, and believe. And it is quite possible that some parts of speakers' intended meaning are less noticeable from a first-person perspective, while being more so from a third-person point of view.

Stephen Schiffer (1992) is clearly concerned with arguing against philosophers' over-intellectualization of normal speakers of natural language. He argues that since the nonphilosopher would not even have access to the *form* of a specification of the property of modes of presentation postulated by a given theory of belief ascription, it beggars belief to suppose that such a person could intend to refer to the property (1992: 513). On this view, we are barred from positing parts into conscious propositional attitudes of which the subject cannot possibly conceive. I rehearse this here to make three related points. First, this is only credible as an account of our access to conscious attitudes. As we have seen, these may only be tenuously related to other mental attitudes. Secondly, intentionalism is in danger of succumbing to another kind of intellectualism, namely, the intellectualism inherent in assumptions of transparency. The apparent fact that people have fairly reliable knowledge of the contents of their mental states is not sufficient to show that there is no gap between the content and the knowledge of that content. Otherwise, our similarly reliable perceptual capacities should point in the same direction, but everyone allows that there is illusion and hallucination in that case (Carruthers 2011: 34). Thirdly, intentionlists who accept mental transparency run the risk of trivializing propositional attitude ascription generally.

That is to say, if a propositional attitude is composed of two items *p* and *q*—one being, e.g., what is said and the other what is implicated—and subjects are generally worse at consciously detecting *q*-type contents than *p*-type contents, attitude attributions by theorists will be systematically impoverished. Arguably, some people are worse than others in consciously detecting some types of propositional attitudes; especially figurative meanings, emotional attitudes, and conversational implicatures. Take figurative meaning interpretation, for example. The ability to interpret and understand metaphorical and ironical utterances can be severely restricted by low IQ, various brain damages and disorders, schizophrenia, and autistic spectrum disorder (Gibbs & Colston 2012: 286-296).

Further, at least if Carruthers (2011, Ch. 10) is right, the empirical evidence suggests that there is no dissociation between other-directed and self-directed mindreading abilities; that is, whenever subjects are cognitively restricted in their capacity to recognize mental states in others they are also restricted in their capacity to recognize those mental states in themselves. Supposing, then, that there is some property F of propositional attitudes such that normal subjects are not very good at finding out about F, transparency-theorists will tend to believe that F is never really a property of propositional attitudes. But there is good reason, especially given the distinction between A-attitudes and B-attitudes, to think there might very well be F-properties. So, transparency amounts to trivializing the contents of mental states by systematically disallowing any F-type property. Now, I have by no means shown that implicit reference to epistemic standards or to modes of presentation must be F-properties of propositional attitudes. But the argument against that possibility was only, as I understood it, that there couldn't be F-properties or, at least, there couldn't be very complicated and unintuitive F-properties which ordinary speakers couldn't possibly conceive of. It follows from the above, however, that epistemic standards and modes of presentation might very well be implicit parts of propositional attitudes, however complex. So, in that respect, they are in the same boat as PRO, location-variables, and the like.

5 Conclusion

I conclude that being a Mad Catter isn't all that bad. Mad Catters believe that it is possible to utter an expression with no phonological properties. They may also believe that only *utterances* of words—not the word-types—have meanings that compose into utterance meanings for the wholes of which they are proper parts. I have argued elsewhere that this is very doubtful (Unnsteinsson 2014), but it is not doubtful because, as Neale claims, there couldn't be such a thing as an utterance of an aphonic. I also

conclude that speakers' lack of conscious awareness that they are using aphonics, and their lack of conscious awareness that they intend to say and/or mean that *p* by uttering something on a given occasion, does not imply that they don't use aphonics or that they don't, on that occasion, say and/or mean that *p*.

Bibliography

Austin, J. L., 1975. *How to Do Things with Words*. Clarendon Press.

Bach, K., 1981. "An Analysis of Self-Deception." *Philosophy and Phenomenological Research*, 41(March):351–370.

———, 2010. "Refraining, Omitting, and Negative Acts." T. O'Connor & C. Sandis (eds.), *A Companion to the Philosophy of Action*, Blackwell, pp. 50–57.

Brand, M., 1971. "The Language of Not Doing." *American Philosophical Quarterly*, 8(1):45–53.

Bromberger, S. & Halle, M., 2000. "The Ontology of Phonology (revised)." N. Burton-Roberts, P. Carr, & G. Docherty (eds.), *Phonological Knowledge*, OUP, pp. 19–37.

Carruthers, P., 2009. "An Architecture for Dual Reasoning." J. Evans & K. Frankish (eds.), *In Two Minds: Dual Processes and Beyond*, Oxford University Press.

———, 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. OUP Oxford.

Carston, R., 2002. *Thoughts and utterances*. Blackwell.

Clarke, R., 2010. "Intentional Omissions." *Noûs*, 44(1):158–177.

———, 2012a. "Absence of Action." *Philosophical Studies*, 158(2):361–376.

———, 2012b. "What is an Omission?" *Philosophical Issues*, 22(1):127–143.

———, 2014. *Omissions: Agency, Metaphysics, and Responsibility*. Oxford University Press.

Collins, J., 2007. "Linguistic Competence Without Knowledge of Language." *Philosophy Compass*, 2(6):880–895.

———, 2008. "Knowledge of Language Redux." *Croatian Journal of Philosophy*, 8(1):3–43.

Davidson, D., 1982. "Paradoxes of rationality." R. Wollehim & J. Hopkins (eds.), *Philosophical Essays on Freud*, CUP, pp. 289–305.

Devitt, M., 2006. *Ignorance of language*. OUP.

Dwyer, S. & Pietroski, P. M., 1996. "Believing in Language." *Philosophy of Science*, 63(3):338–373.

Egan, A., 2008. "Seeing and Believing: Perception, Belief Formation and the Divided Mind." *Philosophical Studies*, 140(1):47–63.

Evans, J. & Frankish, K., 2009. *In Two Minds: Dual Processes and Beyond*. Oxford University Press.

Fischer, J. M. & Ravizza, M., 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Frankish, K., 2004. *Mind and Supermind*. Cambridge University Press.

———, 2016. "Playing Double: Implicit Bias, Dual Levels, and Self-Control." *Implicit Bias and Philosophy Volume I: Metaphysics and Epistemology*, OUP, pp. 23–46.

Gendler, T. S., 2008. "Alief and belief." *The Journal of philosophy*, 105(10):634–663.

Gibbs Jr, R. W. & Colston, H. L., 2012. *Interpreting figurative meaning*. CUP.

Gigerenzer, G., 1999. *Simple Heuristics That Make Us Smart*. Oxford University Press.

Gigerenzer, G. & Regier, T., 1996. "How Do We Tell an Association From a Rule?" *Psychological bulletin*, 119(1):23–26.

Goldman, A., 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Grice, P., 1989. *Studies in the way of words*. HUP.

Hawthorne, J. & Lepore, E., 2011. "On Words." *Journal of Philosophy*, 108(9):447–485.

Hinzen, W. & Sheehan, M., 2015. *The Philosophy of Universal Grammar*. Oxford University Press Uk.

Hunter, D., 2011. "Alienated Belief." *Dialectica*, 65(2):221–240.

Kaplan, D., 1990. "Words." *Proceedings of the Aristotelian Society*, (Supp. vol.) 64:95–119.

———, 2011. "Words on Words." *Journal of Philosophy*, 108(9):504–529.

Langton, R., 2009. *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*. OUP Oxford.

Lepore, E. & Stone, M., 2015. *Imagination and convention: Distinguishing grammar and inference in language*. OUP.

Lewis, D., 1970. "General Semantics." *Synthese*, 22(1-2):18–67.

———, 1982. "Logic for Equivocators." *Noûs*, 16(3):431–441.

Mandelbaum, E., 2016. "Attitude, Inference, Association: On the Propositional Structure of Implicit Bias." *Noûs*, 50(3):629–658.

Miller, J., Unpublished. "The Metaphysics of Words."

Neale, S., 2005. "Pragmatism and binding." Z. G. Szabó (ed.), *Semantics versus pragmatics*, Clarendon, pp. 165–285.

———, 2016. "Silent reference." G. Ostertag (ed.), *Meanings and other things: Themes from the work of Stephen Schiffer*, OUP, pp. 229–344.

Pereplyotchik, D., 2017. *Psychosyntax: The Nature of Grammar and its Place in the Mind*. Springer.

Pickering, M. J. & Clark, A., 2014. "Getting ahead: forward models and their place in cognitive architecture." *Trends in cognitive sciences*, 18(9):451–456.

Pickering, M. J. & Ferreira, V. S., 2008. "Structural priming: A critical review." *Psycho-*

logical bulletin, 134(3):427.

Recanati, F., 2010. *Truth-conditional pragmatics*. OUP.

Schiffer, S., 1992. "Belief ascription." *The Journal of Philosophy*, 89(10):499–521.

Schwitzgebel, E., 2010. "Acting Contrary to Our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief." *Pacific Philosophical Quarterly*, 91(4):531–553.

Searle, J., 1978. "Literal meaning." *Erkenntnis*, 13(1):207–224.

Stanley, J., 2007. *Language in context*. Clarendon.

Unnsteinsson, E., 2014. "Compositionality and sandbag semantics." *Synthese*, 191(14):3329–3350.

———, 2017. "A Gricean Theory of Malaprops." *Mind and Language*, 32(4):446–462.

Wegner, D. M., 2002. *The Illusion of Conscious Will*. MIT Press.

Wilson, T. D., 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Harvard University Press.