Classic Text 33 - Language and Mind: Thought and Meaning

© Andrew Lozovyi VistaCreate

In the previous series of Philosophy of Language Classic Texts we have been chiefly concerned with meaning as a property of linguistic items. Inevitably, this has led us to say or wonder about the mental states of language users because, we assume, that meaning depends on those states in various ways. In particular, in discussing meaning we are led also to consider linguistic competence or understanding. In this study unit we focus on those aspects of mind that are relevant to language, as presented by Devitt & Sterelny in chapter 7 of their *Language and Reality: An Introduction to the Philosophy of Language* (2nd Edition) (1999). We strongly recommend buying or borrowing a copy of their book and reading the chapters in question.[1]

The authors begin with thoughts; the folk idea being that "language expresses thought". **Propositional attitudes**, as philosophers refer to them, are inner states: beliefs, desires, hopes, fears, predicting, wishing, but also fearing, loving, suspecting, expecting *etc*. Although, we have also used the term in the sense of "the relation that a person has with a proposition, such as having an opinion concerning it or responding emotionally to it". (Oxford Languages)

Although we do not have any direct evidence for the existence of propositional attitudes, we use them to explain behaviour. Why did Oscar vote for Trump? We answer in terms of beliefs and desires: because he believed that Trump was dangerous and he believed that a dangerous president was more likely to oppose the Chinese, and that he desired that the Chinese be opposed *etc*. Thoughts, we believe, are behaviour controlling states. (p. 137)

In Classic Text 17 we examined the hypothesis that language is representational, and hence about thoughts. Thoughts are inner representations (or misrepresentations) about the external world and as such they have *content*. The desire to meet Kim Kardashian is different to the desire to meet the most followed Secret Service agent because these two thoughts have different representational contents. Even if Kardasian were the most followed Secret Service agent, known only to herself and her handlers, we would encounter the same problem of opacity again. The two cannot be

[1] We have not devoted a separate study unit to the author's chapter 6 on syntactic structure because, although important, it comprised mostly of linguistics, and so would not be examined as part of the philosophy curriculum.

substituted *salva veritate*, even if they refer to the same individual. Furthermore, the content of the thought is causally linked to behaviour. If someone wished to meet Kardashian, presumably he would line up outside the TV studio stage door. If however, he wished to meet the most followed Secret Service agent, surely he would have done something else. Similarly, if Oscar believed that a dangerous president was undesirable, he would have voted for a different candidate. (p. 137 - 138)

According to the authors, thoughts differ not only in their representational content: the same content may be involved in a belief, a desire, and so on. Each of these feeds into behavioural control in a different way. Having a thought then involves a relation or an "attitude" to its content, such as believing, desiring, hoping, fearing that... *etc*.

> In sum, thoughts are inner states of people (and possibly other things) that have their causal powers partly in virtue of their representational contents and partly in virtue of the relation people have to these contents.

We can use the notation of the logic of relations (Critical Reasoning 14) to ascribe a thought to a person x using a sentence of the form $x\mathrm{V}$ that p or Vx that p, where 'that p' specifies the content and 'V' is the relation in which x must stand to the content. *E.g.* Oscar believes that Trump is dangerous. However, the term 'thought' is ambiguous. It can be used either to refer to a mental state, as above, or it can be used simply to refer to the content. In this way, it is similar to the term 'proposition'. We will have to be sensitive to the context to decide what is meant. (p. 138)

The Language-of-Thought Hypothesis

Attributed to Jerry Fodor (1975) and Gilbert Harman (1973), the language-of-thought hypothesis proposes that thoughts are inner representations similar to the sentences of human languages (*e.g.* English) and hence linguistic in character. Thoughts, in other words, are *mental* sentences composed of *mental* words. Devitt & Sterelny enumerate several motivations in favour of the hypothesis:

1. Thoughts seemingly have the same semantic properties as human languages.

    a. They have referential relations to the world, just as sentences do. Oscar's belief about Trump being dangerous is about Trump in the same way that the English sentence 'Trump is dangerous' is about Trump.
    b. Beliefs and assertions have truth values. They are either true or false. Hopes and desires, on the other hand, are neither true nor false, but, like requests, they do have **compliance** or **satisfaction conditions** under which they would be fulfilled.
    c. Thoughts, like sentences, can stand in inferential relations. Oscar might have arrived at his belief about Trump being dangerous via his belief that all Republicans are dangerous and that Trump is a Republican, therefore Trum is dangerous. According to the authors, "... the **representational** "**content**" of a thought seems to differ only in name from the representational "meaning" of the sentence used to express or communicate that thought".
    This does not require that a thought is syntactically like a sentence because its meaning is not partially dependent by its sentential syntax. Besides which, there are other ways of having meaning. Think of a map, a diagrammatic set of instructions, or even a certain naval flag that historically meant *yellow fever on board*. (p. 138 - 139)

2. Thoughts are similar to sentences, not only in being harbingers of meaning but also having syntax similar to a sentence. One reason for thinking so is c. above. How could Oscar have inferred his belief from the other two? Oscar's thought process in arriving at his belief seems to have been governed by the syllogism of the form: All F's are G's, a is an F, $\therefore a$ is G. Oscar's inference is an instance of this form which has a certain syntactic structure shared with sentences. (p. 139)

3. Another reason for attributing some kinds of syntax to thoughts is that thoughts, like language, are systematic. When humans, and other species, acquire language, they do not learn to understand sentences one by one, but also the elements of sentences and the rules for combining them. This requires that sentences have a syntactic structure. Similarly with thoughts. Humans language users have the capacity to think infinitely many thoughts that may never occur to them, yet they have grasped the concepts encompassed in those thoughts as well as the rules for combining them. This capacity too requires that thoughts have a syntactic structure.

   These reasons militate against the view that thoughts are structurally simple like a naval flag, but they are not so persuasive that thoughts are like maps, diagrams or images; yet they appear to be systematic in a way, at least sometimes. Formal logic provides us with a framework for how an inference like Oscar's might be constructed if the components were represented linguistically. From it's earliest days computer science has used this idea to build machines capable of processing linguistic representations. Indeed computer languages have many of the features of natural languages including syntax and an interpretation or semantics intended for the users. Early **connectionist models**[2] using mathematical representations of a different sort have had some success with problem solving but were a far cry from anything like capturing human inference. Recently the chatbot and virtual assistant ChatGPT, trained on a gigantean corpus of texts, have persuaded some users that it represents a genuine form of **artificial intelligence**[3], even though no one, including its creators, understands its exact mode of representation or inferential procedures. Whether such systems are genuinely intelligent and insightful or merely simulate human intelligence to a greater or lesser degree is probably beyond the scope of the study of language alone. (p. 139 - 140)

4. The authors provide two further reasons in favour of the language-of-thought hypothesis. Firstly, thoughts, like sentences, are abstract. The sentence, "Orson weighs 130kg" tells you only Orson's mass and nothing else. (Classic Text 17) Images, maps and diagrams, however are too rich and ambiguous to capture the content of a thought. A picture of Orson in later life is no more a representation of the belief that Orson was overweight than it is a

[2] These take inspiration from the manner in which information processing occurs in the brain. Processing involves the propagation of activation among simple units (artificial neurons) organized in networks, that is, linked to each other through weighted connections representing synapses or groups thereof. Each unit then transmits its activation level to other units in the network by means of its connections to those units. The activation function, that is, the function that describes how each unit computes its activation based on its inputs, may be a simple linear function, but is more typically non-linear (for instance, a sigmoid function). (McClelland, J. & Cleeremans, A., 2009 edited) The full entry is available here.

[3] Artificial intelligence is defined as, the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages (Oxford Languages)

representation of him being Caucasian, having a receding hairline or displaying any other visible trait. Pictures and thoughts are abstract in the same way. Although pictures may be *associated with* perceptual thoughts, they are not themselves thoughts. (p. 140)

5. Finally, we need to *explain* the contents of thoughts. The thought that Trump is dangerous, and the sentence, 'Trump is dangerous' have a component that means *Trump*. We need to explain how this component contributes to the meaning of the whole. If we are concerned with sentences we look to linguistics to reveal syntactic structure and to logic for how truth conditions depend on structure. Of course, we may lack the finer details, but they could always be filled in later. If the language-of-thought hypothesis is correct then the same approach for language should work for thoughts, because the contents of thoughts are the meaning of *mental* sentences.

   Nor do we seem to have any better explanation. Consider maps, for example. We can see that a map of South Africa clearly represents *Johannesburg being North of Cape Town*, but we have no idea of how it would represent non-spatial relations such as *Felix being a cat* or *Trump being dangerous*; even less so quantifiers, counterfactual conditionals or other complex thoughts. But of course, maps are only intended to represent spatial relations between named features. Nevertheless, the authors do endorse the language-of-thought hypothesis, *i.e.* thought is in a language. (p. 140)

A Public Language of Thought or "Mentalese"

So, what language is thought *in*? Perhaps a person's language of thought is similar in various ways to the public language in which they express their thoughts. Perhaps the two are the *same*. According to this **public language-of-thought hypothesis** a person's language of thought is in their language of speech or signing. Alternatively, a person's thought is in a special mental language or **Mentalise** (měn'tl-ēz'). So, if a person's native language is English, then according to the **Mentalise hypothesis**, they speak by *translating* Mentalese into English and they understand by translating English back into Mentalise. According to Fodor's strong version of the Mentalise hypothesis, Mentalise is both universal *and largely innate*. The authors are sceptical of the strong version of this hypothesis, and instead discuss the moderate version. (p. 140 - 141)

The central question is whether the language of thought is a public language, or some kind of Mentalise. Note that this question is distinct from the previous question of whether we think in a language. This one asks further, whether we think in one language and not another.

One easy to dismiss objection to the public language-of-thought hypothesis is as follows: when we examine the brain of langue speaker or signer, there is nothing that resembles a language. But then what would a language look like? Sentences, unlike pictures are medium independent. (Classic Text 17) Whereas the relation between a picture and what is depicted may be natural or intrinsic, the relation between a name and what is named is arbitrary. Potentially, anything could be used to refer to anything. In Classic Text 25 we saw how very different physical types could be used for one semantic type: a sentence may be physically realised as acoustic vibrations, gestures, inscriptions, the display of flags, electronic impulses, and so on. Therefore, there is nothing preventing sentences being realised in a neutral medium as well. It could be that tokens of thought, like tokens of speech or gestures, are sentences of a public language.

According to the authors, this hypothesis must at lest be qualified: not *all* thinking could be in a public language. Some non-human animals (Classic Text 17) and prelinguistic children undoubtably think, but not in a public language. Similarly, some mature human thought is not in a language either: consider thoughts about music or chess. Thus revised, the hypothesis would be that *most* mature human thought consists in having attitudes of believing, desiring, hoping *etc*. to mental sentences in a public language. (p. 141)

This hypothesis is intuitively appealing because our cognitive capacities appear to be very closely correlated with our linguistic capacities. Indeed the two appear to be developmentally correlated. Furthermore, it is plausible to suppose that our ability to think certain thoughts depends on language. Could we have thoughts about aircraft carriers or nuclear weapons if we had no words or signs for them? Intuitively, speech often *seems* to be "thinking out loud" and thought, it often *seems*, to be "talking to oneself". Consider learning a foreign language. It is not enough to learn the grammar and vocabulary, one must learn to "think in the language". Until one masters this skill, employing a foreign language is a matter of translating back and fourth "in one's head" to one's native language. Or, so it seems introspectively. (p. 141 - 142)

These considerations do not unequivocally support the public language-of-thought hypothesis because the Mentalise hypothesis may well be able to accommodate them too. Perhaps the simultaneous development of language and thought can be accounted for by the same causes, both environmental and neurological. (*Cf*. Critical Reasoning 12 - The unspeakable Experiment) The Mentalise hypothesis could account for our introspective experience of thinking about another language. Before mastering the art of thinking in another language it seems that there are two steps of translation involved: first, translating Mentalise into one's native language, and then retranslating into the foreign language. The first is practiced and unconscious, the second effortful and conscious. When a person masters the art of thinking in the foreign language, the two steps are collapsed into one unconscious one.

According to the authors, there is some introspective and experimental evidence (not cited) that militates against the public hypothesis. How often have we had a thought and yet had trouble expressing it? We say, "the words are on the tip of my tongue". Yet if thought were in the very words we use to express it, why would we sometimes have this difficulty? This phenomenon suggests that either we are experiencing difficulty translating from Mentalise into our native language, or that we are struggling to form the thought, which once formed in our native language can be easily articulated. (p. 142)

Further experiments (again not cited) show that when we read a passage, we tend to remember the "message" rather than the wording. If presented with a sentence that did *not* occur in the passage and asked whether it did appear, we are more likely to say that it *did* if it has much the same meaning as the sentence that was in the passage. One the one hand, this suggests that we have a stored representation in Mentalise but not of the several versions in the language that actually appeared in the passage. On the other hand, perhaps we did store a representation in the language of the text, but that our primary interest was in the message and not the particular words in which it was presented. Alternatively, we may have stored it in a form that was "natural" for us, or one more suitable for later retrieval. (p. 142 - 143)

There are several objections to the public hypothesis that call attention to the fact that public languages are often inexplicit but that the language of thought cannot be. Thus a public sentence "looks like"

(1) John hit the boy.

But the same mental sentence must "look more like"

(2) [$_{S}$[$_{NP}$[$_{N}$John] [$_{VP}$[$_{V}$hit] [$_{NP}$[$_{DET}$the] [$_{N}$boy]]]].

In the notation of linguistics here, S refers to a subject, NP to a noun phrase, N to a noun, VP to a verb phrase, V to a verb and DET to a determiner, in this case the definite article.

(2) above looks like (1) with to introduction of some syntactic notation; however the difference is manifest in that public languages are *ambiguous* whereas the language of thought cannot be. *E.g.* the English sentence, 'Visiting relatives can be boring' is typically ambiguous. However, according to the authors, when a person has a thought that expresses this sentence, the thought cannot be ambiguous if the person's thinking is to proceed in the appropriate rational manner. Executive function can only operate on a thought sentence that makes it explicit whether it means that the visiting of relatives can be boring or whether it is the relatives who visit who are boring.

According to the authors, this objection confuses ambiguity and the significance of explicitness. A certain type in a physical medium, *e.g.* sound, is ambiguous if some of its tokens have one meaning and others another. (See Classic Text 15) They do concede that the objection is right in that a thought expressed on a particular occasion, *i.e.* a mental *token*, is not ambiguous, and neither is the utterance token that expresses it. Ignoring confusion and deliberate puns, any token of 'Visiting relatives can be boring' will have one meaning or another. Ambiguity is neither a property of mental tokens nor the utterance tokens used to express it. The objection rightly points out that explicitness is different. Whereas the utterance token has some syntactic properties that are not explicit, all syntactic properties of Mentalise must be explicit. If this is the case then there can be difference in the physical forms of the tokens, but this does not show that they differ in their syntactic properties. Written tokens of 'Cheap food and wine can be interesting' and 'Cheap (food and wine) can be interesting' differ in form but can be syntactically identical; so too with (1) and (2) above. Therefore, "whether a token has a syntactic property is one thing, whether or not its form makes the property explicit is another. Thoughts differ from utterances in that they must be *syntactically explicit* not in that they must be *syntactically different*". (Our emphasis, p. 143)

Noam Chomsky, sometimes referred to as "the father of modern linguistics", has argued strongly against the public hypothesis and in favour of a cognitivist account of linguistic competence. At the time of writing, Chomsky believed that human language is overwhelmingly rule governed. Linguistic behaviour, in particular, is controlled by rules explicitly represented in the mind of the speaker – rules of which the speaker is cognoscente. These rules dictate which strings of words count as sentences, and what these sentences mean. Most of these rules are *innate*. In other words, we are neurologically hardwired with information about the kinds of rules needed to learn a language. Chomsky refers to this as a **Universal Grammar**, of which we have innate knowledge; however the theory of a universal grammar remains controversial. (Christensen, 2019)

According to Chomsky, speakers understand their native language in terms of knowledge of rules for its employment, but these cannot be *in* the native language. In the same way that a Japanese-Japanese dictionary is of no use unless one already understands Japanese, so the rules telling you how to construct sentences in English and what they mean are of no use unless you already understand English. It follows that a language user must represent those rules in the language of thought *i.e.* Mentalise, not in any public language like English, Japanese, Swahili, Greek *etc*. According to Fodor (1975), Mentalise must therefore be *at least* as rich as any public language whose rules it represents.

According to the authors, the public hypothesis allows for some limited place for Mentalise, much less than Chomsky or Foder foresaw. The role of Mentalese, on the hypothesis, is a relatively impoverished system of representation present in some animals and pre-linguistic children, that is a supplement to the public language representations of human adults. According to the hypothesis, the rich understanding of Mentalese in the previous paragraph stems from the assumption that linguistic competence consists of *knowledge*, and from certain assumptions about the nature of learning. In their nest chapter the authors argue against these assumptions. Although they find some merit in the idea that the *syntax* of the language of thought is partially innate, they cast doubt on the idea that vocabulary could be. (p. 144)

The authors also doubt that any evidence could decisively demonstrate whether the language of thought is a public language or Mentalise. What would be required to for a mental sentence to be a sentence of English, for example? Clearly it must consist of English words and English structure. But what would make a mental word count as an English word? Is there a theoretically interesting answer? Probably not. Instead, the authors decide to treat the language of thought as a distinct Mentalise and focus instead on the more theoretically productive issue of how close a persons Metalise is to their public language. (p. 144 - 145)

According to the following argument, the syntax of an English speaker's Mentalise cannot be very different from that of English. The process of translation, producing and understanding English must preserve meaning, in some sense. In other words, when a thought is expressed by an utterance, the utterance must mean the same as the Mentalise sentence involved in the thought. Similarly, when an utterance is understood, it must be assigned to a Mentalise sentence that means the same. But the meaning of a sentence is a function of its syntax; therefore the syntax of a speaker's Mentalise sentence has to be close enough to that of the English expression if the expression for it is to mean the same.

Without an account of the sense in which a thought matches its expression in meaning, there is quite a lot of room for difference in the syntax of the two tokens. *E.g.* an English sentence in the active voice means the same as the same sentence in the passive voice, despite their syntactic differences. Furthermore, there is a sense in which a literal sentence in English means the same as its Japanese translation, despite their syntactic differences.

According to Fodor, there is another reason for believing that the syntax of a person's Mentalise is very similar to their public language, namely the *speed* of language processing: "the more structural similarity there is between what gets uttered and its internal representation, the less computing the sentence understander will have to do". (*Op. cit.* p. 152) In other words, the more syntactically alike they are, the faster the process of translation from an utterance to the mental sentence that is its interpretation.

Returning to the Problem of Language in Classic Text 17, the task there was to explain "meanings" as properties of utterances that enable them to play a role in explaining behaviour and informing us about the world. How do these utterances differ from birdsong or the "waggle dance" of honey bees? Moreover, what should settle the debate over whether Alex the African Grey Parrot and non-human primates who have been taught to sign, are using language? According to the authors, for an utterance to be in a language, they have to be expressions of thoughts with the rich syntactic structure of our thoughts. English is the expression of thoughts that has just such features. (p. 145)

Contrary to what the authors claim, it seems to us that when Alex the African Grey Parrot did use terms for the 50 or so different objects that he could recognise, including the concepts 'bigger',

'smaller', 'same', and 'different', and the numbers 1 to 6 including 'none' for zero, to express thoughts that meant just what he said. (Classic Text 17) Of course human thoughts and Parrot thoughts likely to be subjectively very different given their very different brains; however we have already discussed multiple realizability in Classic Text 11 as the idea that the same mental property, state, or event may be realised by multiple physical properties, states, or events. Unfortunately the authors could not have been aware of the most important literature about Alex because it postdates the publication of their book. Also unfortunately, the authors devote only a single sentence to "controversy" over whether non-human primates can sign "utterances" that express such thoughts. According to our reading, when Washoe the Chimpanzee signed to one of his caretakers "MY BABY DIED", that is exactly the thought she expressed. (Classic Text 17)

Grice's Theory of Meaning

Herbert Paul Grice first published his theory of meaning in 1957 and included latter developments in book form in 1989. Beginning with the problems of vagueness and ambiguity of the term 'meaning' Grice distinguishes several species of meaning. Firstly, what he called **natural meaning** is causal or a 'reliable sign of'. Thus, 'those clouds mean rain' mean that those clouds *are the cause of*, or *a reliable sign of* rain. At the time, Grice assumed that these were non-semantic. On the other hand, Grice identified two species of **non-natural meaning** or **meaning$_{NN}$**, which are semantic: the *standard*, *literal* or *conventional* meaning of a sign, and *what a speaker means* by a sign on some occasion. Mostly, conventional and speaker meaning coincide, but sometimes they do not. Thus, in the book *The Old Dick* (Morse, 1981), the protagonist describes a thug as 'so primitive that he could regenerate missing limbs'. Literally or conventionally, this sentence is clearly false. However what the protagonist means by the utterance is true, that the thug is extremely stupid. With metaphors like this one the speaker means something different from what his words conventionally mean. Metaphoric meaning is derived from the conventional meaning, but to a degree is independent of it. (p. 146)

There are other situations in which there seems to be speaker meaning without conventional meaning. Consider the likely original development of language. Probably, utterances included noises or gestures were used with communitive intent before there existed a settled system of conventions for using them. According to the authors, communicative effort was, at least, partially successful as a precondition for linguistic conventions. Even now there are situations in which there is speaker intent without conventional meaning. When people without a common language must communicate without an interpreter, somehow they partially succeed via gestures, mimes, body language and contextual cues. (p. 146 - 147)

Speaking metaphorically involves deliberately deviating from conventional meaning, although sometimes this can be accidental. When we say, "You didn't say what you meant", we are usually indicating that there is just such a divergence. A slip of the tongue, on the other hand, may be conventionally ungrammatical, and hence lack conventional truth conditions, yet we have speaker syntax and truth conditions sufficient to allows us to understand the misspoken words. What do we make of the joke about the psychotherapist who tells her client, "A slip of the tongue is when you mean to say one thing, and instead you say your mother".

The authors point out that they have been writing as if the conventional meaning of a sentence and its literal meaning are the same; however there are times when this is not so. A person may have an eccentric **idiolect** (unique use of language or way of speaking) in which the literal meaning of her expression may not be the meaning it has according to linguistic convention. Consider Davidson's

(1986) example of Mrs. Malaprop's "a nice derangement of epitaphs". What she literally means is: "A nice arrangement of epithets", yet her words do not mean this according to any convention.

There is also a distinction to be made between speaker and conventional meaning that applies to pragmatic features of a sentence, particularly to **illocutionary force** *i.e.* a speaker's intention in producing an utterance or the effect a speech act is intended to have by a speaker. Consider:

> I promise you that if you butt your cigarette on my rook again, I'll call the tournament director.
>
> Is it within your capacities to pass that jug of water? My wooden leg has caught fire.

The first sentence is likely to be a threat or a warning. The second is unlikely to be intended as a question about the hearer's abilities as a waiter. Yet both sentences have the conventional forms of a promise and a question, respectively. These cases, known as **indirect speech acts**, can be accommodated by adapting Grice's terminology. We recognise that their conventional force is different from their speaker force, analogous to the difference between conventional and speaker meaning. Moreover, their illocutionary force seems to be part of their meaning in a way that they can be regarded as similarly analogous. (p. 147)

If we suppose that the distinction between conventional and speaker meaning is real, we may wonder which is more basic or prior. According to Grice, it is speaker meaning that must have come first. Some of the examples above reveal that speaker meaning sometimes exists without, or independent of, conventional meaning; however conventional meaning cannot be similarly dethatched. (p. 147 - 148)

We can think of examples of a sentence having conventional without a speaker or writer meaning anything by it. One hundred monkeys tapping away on one hundred typewriters may, now and then, produce a sentence of Shakespear. Even more unlikely, but not impossible, the wind might carve out the words "Trump is dangerous" on a rockface in the Mojave desert. But would such chance words really have meaning? In both cases, there is no matter of fact what the words refer to. According to the authors, whenever a token is of a type that is ambiguous within, or between languages, it is impossible to assign to is a conventional meaning in the absence of speaker meaning, because conventional meaning depends on which convention the speaker had in mind. Of course, even if the token were of an unambiguous type, we could simply decide to assign to it a conventional meaning, but what would be the point? And even if there were a point, such examples would not cast aspersions on Grice's insight, *i.e.* a sentence could not have a conventional meaning that is not derived from *past* regularities in speaker meaning. (p. 148)

According to Grice, a communicative intention is in some way reflexive: a person intends to communicate by means of their audience's recognition of *that very intention*. Consider the following contrast that led Grice to this view. Suppose that Tom wishes to produce in Dick the belief that Dick's lover is having an romantic affair with Harry. Tom might use one of the following two tactics:

1. Tom arranges for Dick to see a photograph of his lover and Harry in compromising circumstances
2. Tom draws a picture of Dick's lover and Harry in such circumstances, and shows it to Dick.

Grice makes the following claims about this scenario. Firstly, he supposes that a photograph has natural meaning but no meaning$_{NN}$; the drawing however does have speaker meaning. Secondly, there is a difference in the role that intentions play in the two cases. Dick may come to the belief that

Tom intends from the photograph without him being aware of Tom's ploy. Maybe Dick believes has stumbled across the photograph quite by chance. Contrast this with the case of the drawing. Unless Dick takes cognoscence of the fact that Tom drew a picture of his lover and Harry, and drawing it with the intention of inducing his intended belief, Dick will not arrive at that belief. If Dick takes Tom to have a different intention, *e.g.* producing a depraved image, Tom will not achieve his aim. Dick must recognise Tom's intention and Tom knows it. (p. 148 - 149)

Grice's remarks on this case are plausible and led him to the following account of speaker meaning:

> 'A meant something by x' is (roughly) equivalent to 'A intended the utterance of x to produce some effect in an audience by means of the recognition of this intention'; and we may add that to ask what a meant is to ask for a specification of the intended effect ... (Grice, 1957 p. 442)

Grice and others revised and complicated his original account in a way that was not psychologically plausible and did not rule out all suggested counter-examples. The authors claim that Grice's discussion of intentions is behaviouristic, *i.e.* to have intentions is simply to be disposed to behave in certain ways. But the behaviourist program does not deal with mental states, if only because they do not necessarily manifest in overt observable behaviour. Besides which, we have already dismissed behaviourism in multiple contexts. If Grice and others' later developments of speaker meaning are correct, then their complex structure of intentions must be part of the unconscious mental life of speakers, which is also not accommodated by behaviourism. (p. 149)

The authors suggest that Grice's later definitions are so exceptionally complex because he saw himself engaged in the task of "conceptual analysis" or the "analysis of ordinary meaning". In other words, his definitions were attempts to analyse the concept of ordinary meaning. Analysis however, must be constructed out of common sense concepts that are familiar to all. And Grice's analysis boils down to just two familiar elements: intention and belief. Secondly, any analysis must be both necessarily true and knowable *a priori*; therefore any analysis must be inviolable to our intuitions about *any* situation, real or consistently imagined "thought experiments". The complexity of Grice's later definitions reflect this difficulty, including the impossibility of covering such an enormous range of potential counter-examples using just a couple of basic elements. (p. 149 - 150)

The authors have a different view of the task of the philosophy of language. From their naturalistic perspective, the most that an analysis could uncover is our implicit folk theory of language, but that is only the beginning. It is an open empirical question as to the quality of that theory. How well does it explain the phenomena involved? It is certainly likely to be incomplete. It could also be dead wrong. If conceptual analysis reveals a distinction between speaker meaning and conventional meaning, then what it reveals is a folk distinction. What matters to the authors is not whether it is a folk distinction, but whether it is theoretically useful, and they think it is.

If analysis shows that it is difficult to explain speaker meaning in terms of familiar notions, this may indicate that our folk intuitions about some alleged counter-examples are misguided, and hence that the folk theory is itself wrong. Alternatively, it may indicate that our folk notions are unable to explain speaker meaning, in which case our folk theory is incomplete. We may have to appeal to an explanation in terms of unfamiliar notions. Such an explanation may appeal to aspects of psychological organization that differ from our intentions and belief, and that are counterintuitive to folk psychology. (p. 150)

The authors have a more grave objection to Grice's approach to speaker meaning. At best it may distinguish communicative acts from other kinds of human behaviour, and perhaps the illocutionary force of such acts. But it tells us nothing about the *content* of such acts, *i.e.* nothing about what distinguishes one speaker meaning from another, beyond perhaps, their illocutionary force. In virtue of what does a speaker mean when he says, "Trump is dangerous" beyond that Trump is dangerous, and not that ducks quack. According to Grice's account, the former and not the latter was the content of the belief that the speaker intended to convey. This accords with the folk notion that an utterance expresses the thought that underlies it, a notion that the authors have endorsed. But this raises a deeper question: In virtue of what was *that* the content underlying the thought? What makes it so that the content was *Trump is dangerous* and not *ducks quack*? Grice and his followers provide no answer. (p. 150 - 151)

Can we provide an answer? If we were to explain a thought's content simply in terms of its direct causal relation to the world it concerns and/or its relation to other thoughts, we would not have a problem. However, our theory of reference borrowing implicitly allows the conventions of public language a role in explaining the contents of thoughts. Given that the authors initially endorsed Grice's program, this potentially leads to a explanatory circle. (p. 151)

Avoiding the Explanatory Circle

The potential explanatory circle arises as follows; the explanations are indented: Suppose that a person has a certain thought.

a) We have endorsed the language-of-thought hypothesis; therefore we identify the content of this thought with the meaning of the sentence involved in the thought.

Suppose now that the person expressed their thought in their public language.

b) According to both Grice's and the folk view, the content of the thought determines what the person means by the sentence they utter.

c) Also according to Grice, the conventional meaning of their sentence in the public language can be explained in terms of regularities in speaker meaning. *I.e.* conventional meaning depends in some way on what people have commonly meant by words of the physical type characterised in the sentence, and on what people have commonly meant by sentences of the same structure. (See Classic Text 25 & 29)

d) Our theory of reference borrowing (Classic Text 25) implicitly appeals to the conventions of public language to explain the meanings of mental sentences. But this step, which is contrary to Grice's priority of speaker meaning leads back to a) and follows.

According to the authors, in order to break the explanatory circle, we must reexamine d) closely. The meaning of a sentence, whether mental or public, is explained by its syntactic structure and the referential properties of the words that occupy that structure. However the meaning of a mental word that depends for its reference on reference borrowing *does* depend on public language convention, but insofar as its meaning depends on independent reference fixing, it *does not* depend on a convention. And of course, no one can borrow reference until that reference has been fixed. It is fixing that explains conventions, but the syntax of mental sentences does not depend on convention.

Instead the meaning of a mental sentence is determined by its causal relations to the world and to other sentences. So while d) is true, it doesn't close the explanatory circle. (p. 151 - 152)

In elaborating this point, the authors presuppose a historical-causal theory of reference fixing. Although such theories have at least one serious difficulty, namely the *qua* problem, the choice between such theories and any other ultimate explanation of reference will not make any difference to this section. (p. 152)

Consider mental words underlying names and natural kind terms, covered by descriptive-causal theories of reference fixing and pure-causal theories of borrowing. (Classic Text 25 & 29) Our contemporaries can have thoughts, including the mental word SOCRATES.[4] Such thoughts would be expressed by inserting the English word 'Socrates' into a causal network grounded in a certain ancient Greek philosopher. For over 2 400 years this network has been established and maintained across languages and alphabets by the convention of using the sound and inscription types for 'Socrates' to refer to the philosopher, and as such uses participate in the convention. Thus, the meaning of SOCTRATES, in its property of designating Socrates by causal chains of the type that constitute this network, is partly explained in terms of the convention. Similarly, most of us have had thoughts about protons in virtue of being linked to them by the network for 'proton' established by the conventional linguistic practices of physicists.

According to the authors, mental words that do not have their meanings in virtue of such conventions are indicated by qualifications that accompany accounts of the dependence of mental meaning on convention. Some people, *e.g.* Socrates' wife Xanthippe, his sons and his parents who named him, thus *fixing* the reference of his name, thought about him without depending on the convention. Furthermore, all those who are involved in *subsequent* groundings of a word or term have thoughts that are dependent for their meaning, not on convention but on direct contact with the appropriate object(s). But it is these regularities of speaker meaning, arising out of convention-independent thoughts, that establish the convention. In this way, speaker meaning is prior to conventional meaning, which is consistent with what Grice believed. In the authors example of Socrates, even the thoughts of those who have not grounded their meaning and who are therefore wholly dependent on convention, *i.e.* all of us, do have meanings that can ultimately be explained in terms of groundings. (p. 152)

Consider next mental words underlying the least basic words that are covered by a description theory of reference fixing – words whose reference cannot be borrowed. The authors suggest, PAEDIATRICIAN, BACHELOR, and HUNTER as examples, though there are thousands of others one can think of. The meanings of such words are determined primarily by their association with other words. This association is a matter a word's "functional role" in the cognitive process of the thinker – not a matter of convention. The meaning of such a word is determined secondarily by whatever explains the meaning of other mental words. According to the authors, convention may play a role in explaining these meanings, but they must ultimately depend on the causal reference fixing of basic words; hence not on convention. (p. 152 - 153)

In Classic Text 29 we looked at other sorts of words that lie somewhere between the fairly basic and the least basic. The reference of such terms depends more on their associated descriptions than the basic ones, but less so than the least basic ones. The authors' theory of mental words underlying these "in-between" words combines elements of theories for the fairly basic ones and for the least

[4] We follow the authors in capitalising mental words to distinguish them from words in a public language.

basic ones. Thus, the dependence of these mental words on convention will be *at most* equal to the limited dependence of the fairly basic ones.

What then determines the semantic structure of mental sentences? What makes a mental sentence one of predication, quantification, *etc.*? According to the authors, this is its functional role, including the sentence's inferential interactions, if any, with other sentences. Thus, TRUMP IS DANGEROUS and PUTIN IS RUTHLESS share the same syntactic structure in virtue of their roles in our lives. Similarly, ALL POLITICIANS ARE CROOKS and ALL CELEBRITIES ARE VAIN share a different structure and role. The first two examples are of predication, the last two, predication and quantification. (p. 153)

The authors are not drawn into the question of what determines the syntax of thought, except to say that it is not determined by the conventions of language. We would expect that the conventional syntax of an utterance is explained by regularities in the way that thoughts with a certain syntax produce utterances of that form. Finally, aspects of illocutionary force that are part of the meaning of a thought are explained by its functional role. The authors propose that what gives a thought the force of a question, statement, threat or promise, is its interactions with various beliefs, desires, intentions and the like. Using certain spoken forms regularly to express a certain illocutionary force may lead that form to be adapted as the conventional one for that force. (p. 153)[5]

In Classic Text 21 we saw how causal theories of reference fair better that description theories. The authors then endorsed Putnam's (1975) slogan, "meanings just ain't in the head", aimed at the view derived from description theories according to which meanings are determined *entirely* by what is in the head. Putman's point is to emphasise that meanings extra-cranial links to reality are also important; however Putman did not claim that no aspect of meaning is determined by what goes on in the head. Some of the examples above clearly are. (p. 153 - 154)

The Origins of Language[6]

Humans, both individually and as a species, are able to think without language (Stix, 2024). However, except for very rare domesticate cases, such as Alex the Parrot and Washoe the Chimpanzee, intelligent animals do think but cannot use language. Even domestic dogs can learn to recognise a surprising number of words, but to what extent they can use them remains controversial. (Pappas, 2023) Although early preconventional thoughts preceded the acquisition of conventions, they need not have been innate. We do however have innate dispositions to respond in different ways to different stimuli. And presumably, these predispositions, together with the stimuli we receive lead us, and some other species, to represent the world in thought. The causal relations amongst these representations, and between the representations and the world, allow these representations to refer in the way that they do. Although these early thoughts may be primitive, they need not be without structure; however without decoding their communication or recording the activity different brain regions associated with them, their structure would be very hard to decern. Perhaps, the

[5] We are surprised that no mention has been made of the **grammatical mood** of verbs whose form indicates modality or various relations to truth or reality. English has five moods: indicative, imperative, interrogative, conditional, and subjunctive but some other languages have more. *E.g.* In English, sentences are in the indicative if they make a statement, but switch to the subjunctive if they express a counter-factual *i.e.* a claim that is false, but could be true.

[6] This section is highly speculative but plausible. Unlike stone tools and fossils, thoughts and words do not survive in the archaeological record; therefore their existence and nature must be inferred from the context in which they are thought to have occurred.

authors speculate, primitive humans, including human babies, had or have a thought structure very much like ours. (p. 154)

According to the authors, mental representations of the world come with theorising about it. Understanding aspects of our environment allows us to anticipate regularities and manipulate or control it to a degree. Perhaps this limited capacity led our early ancestors to express primitive thoughts by means of grunts, gestures, percussion or mimicry, *meaning something* by such actions. Over time these grunts, gestures *etc.* became stereotyped or regimented; hence linguistic conventions. As a result, it became much easier for others to have these primitive thoughts and to express them in conventional ways. Furthermore, a ready way of representing the world became available via these conventional signals, and with this the ability to borrow the capacity to think about things from those who initiated or who are already following the conventions. With the facilitation of primitive thought and the drive to understand the world, including others, more complicated thoughts, and hence more complicated signing and/or speaker meaning and conventions, became possible. If the authors' account is correct, humans, both individuals and as a species, have achieved the near impossible task of lifting themselves up by their own semantic bootstraps. (p. 154 - 155)

Beginning with a language of thought and expanding outwards with the reciprocal acquisition of a related public language lead to a virtuose spiral with feedback going both ways, reinforcing each other. What makes a language public is a conventional form and the regular association of sounds or gestures with speaker or signer meaning. Conventional form, in turn, facilitates speaker or signer meaning, but also introduces mental representations into the language of thought. These representations are causally based on, and have the same meanings as, the sounds and gestures that figure in the conventions. According to the authors, the language of thought became more closely tied to public language, but always remains a little ahead of it. We still have the capacity to think outside of the convention of established public language as evinced by our ability to think new thoughts and express them in new words. We now have the vocabulary to think thoughts that, only a centaury ago, were literally unthinkable. (p. 155)

Perhaps the development of a rich and complex public language in humanoid society was very slow and gradual, but once it became established there was practically no limit on the rate at which it could have expanded. The rate of language acquisition among children is surprisingly fast. According to Feldman (2019)

> ... most children acquire the fundamentals of language effortlessly in the toddler-preschool years, without formal instruction or explicit feedback. By age 5, they have a vocabulary of thousands of words; create sentences with complex grammatical features; differentiate literal from non-literal meanings, such as humor or metaphor; observe the social conventions of conversation; and apply language skills in the service of learning to read. (p. 398)

The linguistic stimuli that children receive include sentences conventionally related to thoughts that are content rich. These in turn make it much easier to conceive of content rich thoughts. Moreover, the language areas of anatomically modern human brains are designed and honed by natural selection to make this task effortless – toddlers and preschoolers learn language at their mother's knee and at play. (p. 155)

Although speaker meanings create the conventional forms of language, it its because we have learned these conventions that we are able to have the rich variety of thoughts that we do, and hence the rich variety of speaker meanings that we do. But the *conception* of convention requires

that some people have thoughts, the contents of which are not full dependent on conventions. Each nascent convention encourages others to have new thoughts, whose contents are partly to be explained by the convention. According to the authors: “Thought contents explain the conventions that explain *other* thought conventions. There is no circle in the explanation”. (p. 155 - 156)

Indicator and Teleological Semantics

Hitherto, including previous chapters, the authors have been working with pure-historical-causal theories of reference fixing, applying those theories to the mental analogues of these terms. Yet these theories have a recalcitrant *qua*-problem, introduced in Classic Text 25. As alternatives, the authors consider two pure-causal theories of the relationship between thought and the world, which is why they had to wait until this chapter. On the Gricean approach, however, the reference fixing of a linguistic term depends on the reference fixing of the mental word it expresses. Therefore, a theory of the one carries over to the other, as we have seen.

When the authors supposed that historical-causal theories of reference attempt to explain how reference is “ultimately” fixed, they had in mind how they could be combined with other theories explaining aspects of reference anchored by ultimate links. An ultimate link in this sense is a direct link between a word and reality; on the other hand, we are by now familiar with two sorts indirect links. Firstly, a person’s word may depend for its reference on other words they associate with it, *i.e.* covered by a description theory or a descriptive-causal theory, but not a pure-causal one. Secondly, it has been suggested that words allegedly covered by historical-causal theories of reference fixing can be borrowed. Therefore a person’s reference of such a word will depend on other people’s reference of the word. If so, its link to its referent will be indirect, via the links of other people to the referent. (p. 156)

According to the authors, those who propose indicator and teleological theories do not supplement their pure-causal theories with description theories or reference borrowing. Instead they believe, implicitly, that a person’s thought stands in direct causal relation to its referent, rather than via other words or other people. However, the rejection of indirect links is not an essential feature of indicator or teleological theories. Furthermore, because such a strong case has been made for reference borrowing and the plausibility of description theories for some terms, the authors are reluctant to jettison such indirect links. Rather, they suggest, that such theories are best constructed as theories of ultimate reference fixing, to which other theories of reference could be supplemented. (p. 156 - 157)

Indicator Semantics, first proposed by Dennis Stampe (1979) and Fred Dretske (1981), proposes that a token represents what tokens of their type are *reliably correlated with*. Thus, the token HORSE refers to horses because there is usually, but not necessarily, a horse present whenever we have a token of that type in mind. The token “carries the information” that a certain situation obtains, in much the same way that tree rings carry information about the age of a tree. Tokens of that type are “reliably correlated” with that situation and hence “indicate” that situation.

But this idea is multiply problematic from the perspective of the meanings of linguistic and mental sentences. Firstly, it seems to be a theory of meaning for perceptual states. Our thoughts and language are not necessarily stimulus dependent. Not all thoughts about horses, say I’D RATHER HAVE A HORSE THAN A PORSCHE, are reliably correlated with horses. I might have such a thought in the shower without any horse present. According to the authors,

> … indicator theories need some internal psychological analogue of reference borrowing. The meaning of stimulus independent thoughts about horses derives it[s] reference from the meaning of perceptual thoughts about horses. Horse musings borrow their reference from horse recognitions.

Secondly, indicator theories would need to be developed to take into account the **compositionality of language** *i.e.* the meaning of a complex [linguistic] expression is determined by the meanings of its constituent expressions and the rules used to combine them. (Wikipedia: Principle of compositionality) And in humans, at least, representation is a complex structure, not simply a mental word. In the example of perceptual representation, the representation would not be HORSE, buts something like THAT IS A HORSE. Because such interactions are the exception rather than the rule, indicator theories need to be developed so that the referential properties of the words involved can be abstracted from these situations. According to the authors, "Other complex symbols can then derive their truth conditions from the referential properties of the of the words they contain". (p. 157)

Thirdly, historical-causal theories and indicator theories give a very different account of reference in Putnam's example of Twin Earth. On historical-causal theories, reference is determined by actual interaction with its referent. Thus, WATER refers to H_2O not XYZ because it is actually grounded in in H_2O not XYZ. However, on indicator theories, reference does not depend on any causal interaction, but on the disposition to interact a certain way. Therefore people on Earth and Twin Earth would be disposed to act in the same way toward H_2O and XYZ because both substances have identical properties, except for the chemical formula of XYZ. Perhaps, the authors suggest, this problem of reference can be solved by stipulating that a word refers to that which it is correlated "under normal conditions". Thus, environments containing H_2O but not XYZ are "normal" for us, and hence the correlation between WATER tokens and H_2O. On Twin Earth however, our doppelgangers live in an environment replete with XYZ but no H_2O, therefore their WATER tokens correlate with XYZ. But what counts as "normal conditions"? It is conceivable that there might be another planet identical to Earth and Twin Earth except that it had oceans and lakes containing variable amounts of H_2O and XYZ depending on the seasons.

Finally, there is the problem of how indicator theories accommodate error? Suppose I "see" a cat crouching in the filed some distance away, but on closer inspection it turns out to be a rather fat pigeon in the foreground. Clearly, I *mis*represented the pigeon by thinking CAT. So, some oversized pigeons seen form a certain distance and perspective can cause tokens of CAT. So what CAT is readily correlated with is mostly cats, but also the odd pigeon or other creatures seen in less than ideal conditions. According to the authors, the problem is that many things that a token of a certain type do not refer to, including some of our doppelgangers, *would* cause a token of that type. (p. 158)

One response to this problem is to invoke the idea of normal or ideal conditions above. The kind of light or perspective for mistaking cats as fat pigeons is not appropriate for fixing the reference of CAT. A token of CAT represents what such tokens are caused by under "normal" circumstances. However, according to the authors, the problem then is to give a naturalistic account of "normality". This has given rise to hybrid theories in which indicator semantics are combined with the idea of biological function. This "teleological" concept, unlike its Aristotelian counterpart, has been made respectable since Darwin. Indicator theorists appeal to Darwinism to show that certain circumstances in which mechanisms evolved are "normal" for the functioning of such mechanisms. Perceptual states typically represent what they indicate in such circumstance. So mistaking the colour of blood for khaki yellow under sodium vapour lamps does not count because our visual apparatus did not evolve under such lighting conditions. Nor indeed did they evolve in the presence of XYZ Earth. (p. 158 - 159)

Unfortunately the problems attendant on the attempt fuse teleology with reliability are overwhelming. Firstly, although we are more likely to be error-free in optimal conditions, our perceptual apparatus has evolved to operate in sub-optimal conditions. For example, most humans have good night vision – our eyes are adapted for seeing by night and by day – so nighttime is a "normal" condition or our vision. Even if we are more likely to make visual mistakes by night, the normal conditions under which our perceptual mechanisms are adapted to operate are not optimal. (p. 159)

Secondly, as Peter Godfrey-Smith (1991) has pointed out that the problem of "false positives" are inimical not only for hybrid theories, but for any indicator theory. Most organisms tend to represent situations in which they detect a predator, food, or indeed any significant entity as being *more often right than wrong*. So what an organism indicates is mostly not what it represents. Natural selection has favoured tolerance of false positives as "normal". Consider hearing a rustling beneath the leaves you are about step on. It could be a small rodent or it could be a venomous snake, or something else. Without knowing which you step aside cautiously only to discover later that it was a field mouse. Being tolerant of minor false positives is an evolutionary stable strategy. Making frequent minor false positives is inconsequential, even prudent, compared to making one serious false negative such as failing to detect a venomous snake in your path. According to the authors, "what nature has selected is a safe mechanism not a certain one".

Jerry Fodor (1987 & 1990) has responded to the problem of error by proposing a different sort of indicator theory without an appeal to teleology. According to Fodor, it is a law that horses cause HORSE tokens; *i.e.* there is a lawful correlation between HORSE tokens and horses. The problem is that there is also a law that the odd cow or zebra also cause HORSE tokens. Fodor's solution is that HORSE *does* refer to horses, and only to horses, because the causal relation between horses and HORSE tokens exists; however the relation between the odd cow or zebra and HORSE tokens is *asymmetrically dependent* on the relation to horses. (p. 159)

The authors, however raise the following objection: Fodor's account relies on some paradigm cases of A referring to as such that as are the cause of A, and bs also causing As, and that it is *obvious that* b*s causing* A*s is asymmetrically dependent* on as causing of A. But the asymmetric dependency that Fodor claims is the case for HORSE is far from obvious. At first blush, it seems that the law that obtains is that horses, the odd cow or zebra cause HORSE tokens. In other words that creatures that typically or occasionally have a "horsey" look, cause HORSE tokens. That is why horses, the odd cow or zebra and even certain billboards cause them. So HORSE refers to things that look horsey; hence someone who thinks HORSE at the sight of the occasional cow or zebra is not misrepresenting them at all. So, according to the authors, Fodor is stipulating, or hoping for, the existence of some very complex causal dependencies, without ruling out a much simpler view of these relationships that fit just as well, or better, with the empirical facts. (p. 159 - 160)

An alternative to indicator theory developed by Ruth Millikan (1984), David Papineau (1984 & 1987) and Karen Neander (1995) relies entirely on teleology in explaining representation by biological function alone. The earlier hybrid theory appealed to function in order to identify the circumstances that fix reference. Thus WATER means water, because under "normal" circumstances, an agent thinks "WATER!" only in the presence of water. The appeal to function is to cash out what those circumstances are, *i.e.* the circumstances similar to those which caused the ability of the token WATER to evolve in our species. On this hybrid view, meaning still depends on indication, but according to the authors, a full-blown teleological theory appeals to function to explain the very content of representational states. Thus, the mental state of humans who cautiously sidestep a rustling sound in the leaflitter is about venomous snakes because its function is to adapt human

behaviour to the presence of venomous snakes in their environment. Biological function, in turn, is explained by the history of selection. The ancestors of humans who cautiously sidestepped rustling sounds in the leaflitter were fitter than those who did not, and fitter because they had a better chance of avoiding dangerous encounters with venomous snakes.

Teleological theories of representation, like indicator theories, are still saddled with the problem of compositionality; however in two other respects they are an improvement. Making a false negative, such as failing to take evasive action when a venomous snake *is* present is far more costly than making several a false positives by evading a non-existent threat. Teleological theories of representation do not have the same problems of error as indicator theories, though according to the authors, they have others. Secondly, the biological function of any structure or system depends on its selective history; therefore like historical theories, teleological theories have no difficulties with Twin-Earth examples. Because XYZ has never been part of the selection of any mental structures on Earth, none of them is an adaptation to XYZ. That we have the concept WATER is hardwired into us by natural selection as a function of H_2O, not XYZ. (p. 160)

However, the teleological theory of representation seems to face a massive problem a *viz.* representations involved in thought. For, what is the biological function of most of human thought? Our capacity to cautiously avoid certain potentially dangerous stimuli seems to be inbuilt by natural selection, so that now it is a feature of all humans. It is very easy to teach children to be afraid of snakes because they seem to be primed for it; however it is fiendishly difficult to teach them about the proportionately much greater dangers of motor vehicles because motor vehicles did not feature in the environment in which our species became adapted.

The authors go out on a limb here. They acknowledge that perhaps a few thoughts have biological functions which specify their meanings, and perhaps those meanings, in turn, explain the meaning of sentences which express those thoughts. They also acknowledge that the size and complexity of the human brain are good evidence for taking it to be an adaptation of some kind, shaped by selection as a behavioural control system. However, to paraphrase them: even setting aside worries about our ignorance of the details of human evolutionary history, at most only a fraction of our thoughts could be part of our biological heritage the way that VENEMOUS SNAKE thoughts are part of our biological heritage… *So surely, most thoughts and sentences have no biological function*. (Our emphasis) (p. 161)

There are two possible rejoinders to this objection. The simplest one, owing to Papineau, is that the learning process is a process of selection, similar enough to natural selection to give beliefs and desires biological functions. The authors however argue that, in addition, the functions would have to explain their meaning. We disagree. Many biological functions have no obvious meaning. Think of the contraction of the gall bladder.

The second rejoinder relies on the distinction between mental states and the mechanisms that produce them. Some species camouflage themselves by taking on the colour pattern of their environment. Suppose that a children's party hat sinks to the bottom of the seafloor and that an octopus comes to rest on it, matching its own colour pattern to the hat. According to Millikan, in such a case, the colour pattern of the octopus has the function of matching the colour of the party hat, even though, in all probability, no octopus has ever had such colouration. A unique state can thus still have a biological function because it was produced by more general mechanisms which were selected to produce particular states of that general type. But according to the authors,

> … even if we accept that thoughts do have functions of this sort, it remains to be shown that those functions explain the content of thoughts. We need to be shown that the function of a desire is to represent the particular state that would satisfy it, and that the function of a belief is to represent the particular state that would make it true… (p. 161 - 162)

The authors are drawn to a less ambitious use of teleology to explain meaning. Instead of supposing that biological functions determine the contents of *thoughts*, they suppose that they determine the contents of more basic representational states, *i.e. perceptions*. "Perceiving a rabbit as a rabbit is a matter of being in a state with the biological function of representing a rabbit." This is an advance on the historical-causal theory of reference fixing, which was saddled with a serious *qua*-problem. The question, in virtue of what is a particular grounding of 'rabbit' a grounding in rabbits, rather than say, mammals or vertebrates or other taxa. The present theory proposes and answer. The grounding is in rabbits because it involves a perceptual state that has a function of representing rabbits. Thus the teleological theory becomes an essential part of the theory of groundings, incorporating teleology into the historical theory of reference fixing. (p. 162)

Task

How likely do you think the Language-of-Thought hypothesis is including the existence of Mentalise.

What evidence is there for the existence of a Universal Grammar?

What do you make of the authors' account of the Origins of Language?

Do you think that teleological theories of representation go far enough in explaining meaning?

Feedback

We have provided what we think is a fair exposition of these hypotheses but we have not put them to the test. At the very least we would like to see them corroborated or confirmed by evidence outside of philosophy. At best, if they are to be scientific theories, they should be falsifiable and repeatedly fail to be falsified if they are successful theories. All four are plausible and help to make sense of linguistic problems in the context in which they arise, but we cannot step out from language and wonder extra-linguistically how things would be if they were false. Unfortunately the authors' account of the Origins of Language is pure speculation and we will never know how or why language appeared and evolved among humans. By the time writing was invented, several times over, globally between 3 400 BC and 1 000 BC, we find language already fully formed. We are most impressed by the teleological theory of perception because it seems to solve the *qua*-problem while offering an account of ultimate reference fixing. Furthermore, it incorporates teleology into the historical theory of reference fixing, which is a noteworthy improvement.

References

CHRISTENSEN, C. (2019) Arguments for and against the Idea of Universal Grammar. *Leviathan: Interdisciplinary Journal in English* **4**: 12-28. doi.org/10.7146/lev.v0i4.112677

DAVIDSON, D. (1986) "A nice derangement of epitaphs". **In** Lepore, E. Ed. 1986: 433-46 *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Blackwell: Oxford

DEVITT, M. & STERELNY, K. (1999) *Language and Reality: An Introduction to the Philosophy of Language* (2nd Edition). Blackwell Publishers Ltd: Oxford

DRETSKE, F. (1981) *Knowledge and the Flow of Information*. MIT Pres: Cambridge, Mass.

FELDMAN, F. (2019) How young children learn language and speech: Implications of theory and evidence for clinical pediatric practice. *Pediatric Review* **40**(8): 398-411 doi: 10.1542/pir.2017-0325

FODOR, J. (1975) *The Language of Thought*. Thomas Y. Crowell: New York

FODOR, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge, Mass.

FODOR, J. (1990) *A Theory of Content and Other Essays*. MIT Press, Cambridge, Mass.

GODFREY-SMITH, P. (1991) Signal. Detection, Action. *Journal of Philosophy* **88**: 707-22

GRICE, P. (1957) Meaning. *Philosophical Review* **66**: 377-88

GRICE, P. (1989) *Studies in the Way of Words*. Harvard University Press: Cambridge, Mass.

HARMAN, G. (1973) *Thought*. Princeton University Press: Princeton

MCCLELLAND, J. & CLEEREMANS, A. (2009) Connectionist Models. **In** Byrne, T., Cleeremans, A. & Wilken, P. Eds. *The Oxford Companion to Consciousness*. Oxford University Press: New York https://philosophy.org.za/uploads_other/McClelland_Cleeremans_2009.pdf

MILLIKAN, R. (1984) *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press: Cambridge, Mass.

MORSE, A. (1981) *The Old Dick*. Avon Books

NEANDER, K. (1995) Misrepresenting and Malfunctioning. *Philosophical Studies* **79**: 109-41

PAPINEAU, D. (1984) Representation and Explanation. *Philosophy of Science* **51**: 550-72

PAPINEAU, D. (1987) *Reality and Representation*. Blackwell: Oxford

PAPPAS, S. (2023) Can Dogs Use Language? The “button dogs” of TikTok seem to be learning human words. What’s really going on? *Scientific American* Newsletter 22 August 2023

PUTNAM, H. (1975) Mind, Language and Reality: Philosophical Papers Vol. **2**, p. 223- 27. Cambridge University Press: Cambridge

STAMPE, D. (1979) Towards a Causal Theory of Linguistic Representation. **In** *Contemporary Perspectives in the Philosophy of Language*, eds. French, P., Uehling Jr., T. & Wettstein, H. University of Minnesota Press: p. 81 -102

STIX, G. (2024) You Don’t Need Words to Think. *Scientific American Newsletter* 17 Oct 2024