New Ideas in Psychology 27 (2009) 163–183

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Meaning and mind: Wittgenstein's relevance for the 'Does Language Shape Thought?' debate

Diane Proudfoot*

University of Canterbury, Department of Philosophy and Religious Studies, Private Bag 4800, Christchurch, New Zealand

ARTICLE INFO

Article history:
Available online 11 June 2008

Keywords:
Consciousness
Language
Linguistic determinism
Meaning
Pinker
Representation
Thought
Wittgenstein

ABSTRACT

This paper explores the relevance of Wittgenstein's philosophical psychology for the two major contemporary approaches to the relation between language and cognition. As Pinker describes it, on the 'Standard Social Science Model' language is 'an insidious shaper of thought'. According to Pinker's own widely–shared alternative view, 'Language is the magnificent faculty that we use to get thoughts from one head to another'. I investigate Wittgenstein's powerful challenges to the hypothesis that language is a device for communicating independently constituted (or individuated) thoughts. I argue that Wittgenstein offers instead a subtle version of the thesis that language determines thought.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The 'Does language shape thought?' debate has revived recently, not only in the social sciences, but also in philosophy (e.g., Bermúdez, 2003; Carruthers, 2002; Chadha, 2007). In this paper I explore, through a discussion of Steven Pinker's views, Wittgenstein's relevance for the debate. Pinker's work is foundational in this area and Wittgenstein provides fundamental criticisms of Pinker's model of the relation between language and thought. In its place, Wittgenstein offers a subtle version of the thesis that language determines thought.

Pinker's claims are explicitly philosophical. The experimental and ethnographic evidence he cites is aimed at theses in philosophical psychology, and frequently his targets are major Western

* Tel.: +64 3 3642781; fax: +64 3 3642889.
E-mail address: diane.proudfoot@canterbury.ac.nz

0732-118X/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.newideapsych.2008.04.012

philosophers. In Pinker's view, understanding the relation between language and thought requires rejecting the theses expressed in two famous quotations, from Locke and Wittgenstein respectively:

> Let us then suppose the mind to be, as we say, white paper void of all characters, without any ideas. How comes it to be furnished? Whence comes it by that vast store which the busy and boundless fancy of man has painted on it with an almost endless variety? Whence has it all the materials of reason and knowledge? To this I answer, in one word, from EXPERIENCE. (Locke, quoted by Pinker, 2003, p. 5)
>
> The limits of my language mean the limits of my world. (Wittgenstein, quoted by Pinker, 2007, p. 134)

According to Pinker, experience is not sufficient for language-acquisition. Rather, language is an 'instinct', dependent upon innate psychological equipment—'a biological adaptation to communicate information' (1995, pp. 18–19). Language is not necessary for concept-acquisition; nor does it 'pervad[e] thought' (p. 17). Rather 'thoughts are merely clothed in words whenever we need to communicate them to a listener' (p. 56).

I argue (Section 2) that in fact Pinker and Locke make similar philosophical claims concerning the relation between language and thought—claims that Wittgenstein strenuously attacked. Both adopt what Wittgenstein idiosyncratically called the 'accompanying picture', in which thought has the primary role: for example, according to the accompanying picture, the meaning of linguistic expressions is determined by internal mental processes and representations.[1] Wittgenstein's powerful objections to this model (Section 3) challenge Pinker's claims, despite his scientific sophistication. As Pinker implies by his quotation above from Wittgenstein, Wittgenstein's own view is that language is primary: mental content is determined by the use of linguistic expressions (Section 4). However, Pinker's various criticisms of theories that prioritize language over thought fail to work against Wittgenstein's approach (Section 5).

2. Pinker and Locke: The same underlying model

There are striking similarities between Pinker's and Locke's fundamental conceptions of language, despite their very different metaphysics of the mind and the greater experimental and logical resources for the investigation of speech that are available to Pinker.

First, the function of language. According to Locke, language is designed (by God) for both the 'communication of thoughts' and to 'record' thoughts (to aid memory). Language enables humans to 'profit' from the ideas of others, for '[t]he comfort and advantage of society' (Locke, 1690/1959, III, ii, §1–2).[2] Speech is 'the common conduit, whereby the improvements of knowledge are conveyed from one man and one generation to another' (III, ii, §1). For Pinker too, language is engineered for communication and, as for Locke, can be used for data storage: 'Aside from its use as a medium of communication, language can be pressed into service as one of the media used by the brain for storing … information' (2003, p. 209; see also Pinker, 2007, p. 129). '[L]anguage is the conduit through which people share their thoughts and intentions and thereby acquire the knowledge, customs, and values of those around them'; it 'allows know-how to be shared at low cost' (2003, pp. 209, 238).

Second, the basic action of language. For Locke, language translates sounds or marks into ideas and vice versa: 'words … stand for nothing but *the ideas in the mind of him that uses them* … When a man speaks to another, it is that he may be understood: and the end of speech is, that those sounds, as marks, may make known his ideas to the hearer' (1690/1959, III, ii, §2). In language 'the thoughts of men's minds [are] conveyed one to another' (III, i, §1); words must 'excite in the hearer exactly the same idea they stand for in the mind of the speaker' (III, ix, §6).

[1] In what follows, I treat 'internal representation' as a place-holder—e.g., for a quasi-linguistic symbol-structure, quasi-pictorial representation, quasi-holographic representation, analogue mental model, or distributed representation with no compositional semantics.

[2] All references to Locke's *Essay* are to A.C. Fraser's annotated edition of the First Edition (New York: Dover, 1959).

According to Pinker, too, language is a translation device. 'Knowing a language is … knowing how to translate mentalese into strings of words and vice versa' (1995, p. 82). (Mentalese is the 'language of thought'.) A word is 'available to convert meaning to sound by any person when the person speaks, and sound to meaning by any person when the person listens, according to the same code' (p. 152). Speakers and hearers 'decode' each other (p. 365). In this way language 'convey[s] a concept from mind to mind' (p. 84); it is 'the magnificent faculty that we use to get thoughts from one head to another' (2003, p. 208). Pinker's picture of language is in this respect extraordinarily Lockean: he says, 'Simply by making noises with our mouths, we can reliably cause precise new combinations of ideas to arise in each other's minds' (1995, p. 15).[3]

Third, language-learning. According to Locke, when children learn words they make 'a constant connexion between the sound and the idea'; this connexion is not natural but 'arbitrary' (1690/1959, III, ii, §7–8). In Pinker's view also, when learning language children 'mentally link the words to the concepts' (2003, pp. 208–9). He says, '[E]very English speaker has undergone an identical act of rote learning in childhood that links the sound to the meaning' and '[s]ince a word is a pure symbol, the relation between its sound and its meaning is utterly arbitrary' (1995, pp. 84, 152).

Pinker criticises Locke's 'associationism' on the ground that language-acquisition, for example, is impossible without innate psychological equipment. Yet even here there are similarities. It is plain that Locke's conception of language has a nativist element. For Locke, as for Pinker, language involves a 'faculty' that 'brutes' lack (1690/1959, II, xi, §10). In Locke's view, '[m]an … had by nature his organs so fashioned, as to be fit to frame articulate sounds' (III, i, §1). In the paragraph Pinker quotes, Locke distinguished 'our understandings' from 'the materials of thinking'. He said, 'These two ['understandings' and 'materials'] are the fountains of knowledge, from whence all the ideas we have', or can naturally have, do spring' (II, i, §2). Locke's view is that concept-acquisition (and hence language-acquisition) depends on human nature and that the mind has cognitive mechanisms independent of the information it acquires from experience.[4]

For both Locke and Pinker, in each of the respects above—the function and basic action of language, and language-learning—thought is prior to language. Cognition occurs prior to the development of language, at the level both of the individual human and the species. 'Ideas' determine—and may even constitute—the meaning of natural-language expressions. In language-acquisition, the meaning of a word is ultimately individuated, not by means of other words, but by internal representations.[5] In short, *language does not shape thought in any important respect*.

2.1. Wittgenstein on the 'accompanying Picture

As we have just seen, both Locke and Pinker subscribe to the following theses: (1) while speaking or listening the language-user performs certain internal mental operations (for Pinker these are computational and most take place in mentalese); and (2) what the language-user means or understands is determined by these mental operations. In short, *thought* gives words meaning (both meaningfulness and a particular content). Wittgenstein called this the 'accompanying picture' of the relation between language and thought (1989b, p. 286). I shall refer to it as the *AP model* of the relation between language and thought.

The core of the AP model is to 'think of meaning or thinking as a peculiar *mental activity*' accompanying (say) speaking (Wittgenstein, 1965, p. 39). On this model [t]he action of language consists of two parts; an inorganic part, the handling of signs, and an organic part, which we may call understanding these signs, meaning them, interpreting them, thinking. (p. 3)

[3] Pinker states, 'Our complicated ideas are built out of simpler ones, and the meaning of the whole is determined by the meanings of the parts' (1999, p. 564). This is exactly Locke's view. Pinker also claims that 'the mind owes its power to its syntactic, compositional, combinatorial abilities' (1999, p. 564); on 'composition' (of ideas) Locke would have agreed.

[4] Pinker notes that 'Locke alluded to something called "the understanding"' (2003, p. 34), but he overlooks Locke's implicit nativism.

[5] It appears, from his most recent book (Pinker, 2007), that Pinker has become aware of some of the problems for these views about meaning. See Section 3.5.

The proponent of the model, Wittgenstein says, argues as follows:

> Without a sense, or without the thought, a proposition would be an utterly dead and trivial thing. And further it seems clear that no adding of … signs can make the proposition live. And the conclusion which one draws from this is that what must be added to the dead signs in order to make a live proposition is something … with properties different from all mere signs. (p. 4)

A classic illustration of the AP model is the distinction, found in both Locke (1690/1959, III, i, §1–2) and Pinker (2003, p. 60), between a human being's and a parrot's grammatically correct vocalisations. According to this model, the difference between the two is that the parrot lacks the accompanying internal mental states necessary for language-use. Wittgenstein remarks:

> You distinguish between talking mechanically 'like a parrot' and 'talking with thought'. Then you try 'thinking accompanies …'. This won't do; It leads … to questions like 'How long did the thought take as compared with the sentence?' (1989b, p. 245)

This question is absurd because *meaning* does not have a *duration* (even a vague one). For Wittgenstein, the absurdity shows that the AP model's interpretation of the distinction between the human's and the parrot's 'speech' is wrong-headed.

In the next section I discuss several powerful objections, arising from Wittgenstein's remarks, to the AP model. For brevity's sake, I focus on Wittgenstein's discussion of meaning.[6] I omit his criticisms of the AP model's related accounts of communication and language-acquisition—in effect, the first and third of the similarities (in Section 2) between Locke and Pinker. (These criticisms include Wittgenstein's attack on the possibilities of a 'private' language and of language-learning by ostensive definition.)

3. Wittgenstein's attack on the AP model's account of meaning

Wittgenstein recognized that the AP model may seem intuitive (perhaps this is due in part to our philosophical inheritance from the 17th century). We are tempted to say:

> [E]ven if words do not stand for or represent things, cannot thought do so? Is not this the peculiar property of mental phenomena? Is there not representation 'in the mind'? (Wittgenstein, 1980a, p. 45)

Wittgenstein claimed, however, that '[t]his suggestion is a pernicious mistake' (p. 45). His aim is to overturn the AP model; only if he succeeds in this can his own account of the relation between language and thought be properly understood—and even appear inevitable.

To undermine the AP model's account of meaning, numerous arguments from different directions are required. Wittgenstein asks three crucial questions of the model. Where is (the hypothesized internal process of) meaning *located* (Section 3.1)? What explains *intentionality* (3.2)? Can meaning be *determinate* (3.3)? In each case, in Wittgenstein's view, the AP model lacks a satisfactory answer. These and other difficulties (3.4) justify our rejecting this account of meaning. In fact, even Pinker appears to reject parts of it in his latest book (3.5).

3.1. Where is meaning located?

On the AP model, meaning is an internal mental process. Wittgenstein asks: *where* is this process—in or outside consciousness? He argues that neither possibility is acceptable; it follows that, contrary to the AP model, meaning is not an internal mental process.

When I speak, am I *conscious* of meaning something? According to Wittgenstein, in practice we may simply not find in consciousness the process of 'meaning' hypothesized by the AP model. I may not experience anything different when meaning my words in one way rather than another (1988a,

[6] Wittgenstein often uses 'mean', 'refer', 'represent', and so on (and their cognates) in an everyday sense, to address a wide target, and where the technical distinctions available in contemporary philosophy are irrelevant. I shall do likewise.

§228, 230), even when a 'whole course of thought' is tied up with meaning them in that way (§230). He remarked:

> Is the hearing or thinking of a word in this or that meaning a *genuine experience*?—How is that to be judged?—What speaks against it? Well, that one cannot discover any *content* for this experience. It's as if one were expressing an experience, but then could not think what the experience really was. (§105)

Wittgenstein introduced the notion of *meaning-blindness* as 'a way of getting rid of … the "accompanying" picture' (1989b, p. 286). The 'meaning-blind' speaker has no 'experiences' of meaning (in contemporary vocabulary, the meaning-blind speaker is a meaning 'zombie'). He or she is 'able to take notice of' uttered or written signs by recognizing only 'the pattern of the sign' (Wittgenstein, 1970, §145) and to manipulate the signs merely 'by means of rules, tables' (§145). The meaning-blind speak 'a language in whose use the impression made on us by the signs played no part; in which there was no such thing as understanding, in the sense of such an impression' (§145). We can suppose, Wittgenstein suggested, a whole community of meaning-blind people. Despite their lacking the 'impressions' of words, they use intentional constructions (such as 'When I heard the word "bank" it meant… to me' (§530)) exactly as we do. For example, the meaning-blind person can say 'Now I've got it!' (1988a, §206) to indicate that she now understands the meaning of an expression. (There are striking parallels between Wittgenstein's remarks on meaning-blindness and Searle's Chinese room argument: see Proudfoot, 2002.)

What are we to say about the meaning-blind speaker? Is she really speaking a language (that is, *meaning* and *understanding* words) or is she merely 'parroting' words? According to Wittgenstein, the fact that the meaning-blind speaker lacks 'impressions' of meaning is beside the point: 'Don't all *experiences* of understanding get covered up by the use, by the *practice* of the language-game?' (1988a, §184). What matters to language is *use*:

> the experience of meaning seems to have no importance in the *use* of language. And so … it looks as if the meaning-blind could not lose much. (§202)

In Wittgenstein's view, the 'meaning-blind' speaker knows the meanings of words, just as you or I. 'Meaning-blindness' is a solecism: 'And now the question arises whether … it wouldn't be totally misleading to speak of … meaning-blindness (§189).

But what if, when *I* speak, I am conscious of something that I might choose to describe as a process of 'meaning'? Wittgenstein said that this would tell us nothing about the nature of meaning or understanding:

> Introspection can never lead to a definition. It can only lead to a psychological statement about the introspector. If, e.g., someone says: 'I believe that when I hear a word that I understand I always feel something that I don't feel when I don't understand the word'—that is a statement about *his* peculiar experiences. (1988a, §212)

Wittgenstein concluded, from considerations such as the above, that the 'psychological processes which are found by experience to accompany sentences are of no interest' in a philosophical inquiry into meaning and thinking (1974, p. 45):

> And here it merely needs to be noticed that 'thoughtless' talk and other talk do indeed sometimes differ as regards what goes on in the talker, his images, sensations and so on while he is talking, but that this accompaniment does not constitute the thinking, and the lack of it is not enough to constitute 'thoughtlessness'. (1983, p. 81)

If Wittgenstein is correct and the internal process of meaning (postulated by the AP model) is not experienced—that is, it does not take place in consciousness—then it must take place at a *non-conscious* level. But prima facie the language-user has no cognitive access to non-conscious processes (if not in principle then at least in practice). The consequence is that I (for example) could never know what I mean, or indeed that I mean anything at all. Also, I could never know what other people mean. This outcome is extremely counter-intuitive. If, on the other hand, we accept the common-sense assumption that I *do* know (or at least have justified beliefs about) what I and others mean, this cannot

be in virtue of some internal process of 'meaning' of which I am unaware. There is simply no role for a non-conscious process of 'meaning' to play. Inferences, as part of a psychological theory, to such a process would not help us to understand meaning. (This is not to reject scientific investigation of language-use; see Section 3.4.)

There is an immediate rejoinder to the argument in the last paragraph. Pinker (1999, p. 135) distinguishes 'access to information' from 'sentience'. (This is Block's (1995, 2007) distinction between access consciousness and phenomenal consciousness.) Perhaps Wittgenstein's meaning-blind speaker has cognitive access to, despite no phenomenal awareness of, the internal process of 'meaning'. However, the question now is: to be access-conscious of some internal process of meaning, need I have some phenomenal indicator—to use Wittgenstein's term, an impression—of the occurrence of this process? If so, Wittgenstein's arguments in this section bite again. On the other hand, perhaps I 'just know' what I mean, in the absence of any phenomenal clue. But why take this 'just knowing' as a sign of an internal non-conscious process of meaning—especially when, as Wittgenstein has argued, positing a conscious process of meaning has no explanatory value?[7]

3.2. What explains intentionality?

How are we to explain the intentionality (meaningfulness) of a word or thought? The Lockean version of the AP model postulates 'immaterial mental mechanisms'—processes that 'take place in a queer kind of medium' (Wittgenstein, 1965, p. 3). Wittgenstein claimed that this is no explanation at all: 'The mental act seems to perform in a miraculous way what could not be performed by any act of manipulating symbols.' (p. 42). The question then is: can a contemporary physicalist version of the AP model, such as Pinker's, offer more?

To illustrate the difficulty posed by intentionality, Wittgenstein suggested that we consider, in place of a hypothetical internal act of wishing, an external observable act:

> Let's say the wish for this table to be a little higher is the act of my holding my hand above the table at the height I wish it to be. Now comes the objection: 'The hand above the table can't be the wish: it doesn't express that the table is to be higher; it is where it is and the table is where it is. And *whatever other* gesture I made it wouldn't make any difference. (1974, pp. 148–149)

Viewed solely as a physical act, holding one's hand above the table has no meaning. In Wittgenstein's words, the intentionality of the act 'is something which is absent from the phenomenon as such' (p. 143). But how, then, is an *internal* physical act or process to have meaning?

Pinker's answer is the *computational theory of mind* (plus causal and inferential-role theories of mental representation (1999, pp. 80–81; see Section 3.3)). Unlike the act of holding my hand above the table, the (hypothesized) processes of meaning and understanding are *computational operations upon internal representations*. 'Language understanding proceeds from raw sound up through representation of syllables, words, and phrases, to an understanding of the content of the message' (p. 139). The 'highest' level of mental representation is 'mentalese, the language of thought in which our conceptual knowledge is couched'; 'Mentalese is the medium in which content or gist is captured' (p. 90).

Pinker's answer, however, is open to the question asked in Section 3.1. Are these supposed internal representations and operations *in consciousness*? To address this question, Wittgenstein used a simple

[7] Pinker defends the computational theory of mind (see Section 3.2) against Searle's Chinese room argument. He remarks:

> My own view is that Searle is merely exploring facts about the English word *understand*. … We can look for another word, or agree to use the old one in a technical sense; who cares? … Science, after all, is about the principles that make things work, not which things are "really" examples of a familiar word. (1999, p. 95)

So do Wittgenstein's objections to the AP model merely explore facts about the English word 'mean'? On the contrary, Wittgenstein would argue that the hypotheses Pinker embraces fail to explain how meaning works. Replacing the word 'mean' (or changing its sense) would not answer Wittgenstein's criticisms.

thought experiment (his actual example is reading 1972, §156 ff.). *A* is a fluent speaker of Mandarin who speaks briefly about an object *o*; *B* is a monolingual speaker of English who has memorized the same words of Mandarin, so as to appear more knowledgeable than in fact he is. Let us suppose that, for any internal representation or computation of which *A* is conscious while talking about *o*, we eliminate it in *A*, leaving everything else unchanged, and induce it in *B*. In *A* we induce the 'characteristic sensations [of] reciting something one has learnt by heart' (§159). (Adapting Wittgenstein's thought experiment to Pinker, we are taking mentalese out of a language-user's consciousness and giving it to a human 'parrot'.)

In this experiment, does *A* no longer refer to *o*? Not according to Wittgenstein: *A* continues to refer to *o*, precisely because the meaning-blind speaker can refer to *o* in the absence of any associated item in consciousness. (This is because it is *use* that matters; see Section 3.1.)

Does *B* now refer to *o*? In effect Wittgenstein asked: why think this? In his view, introducing internal representations and operations 'only adds one phenomenon to another' (1979, p. 54):

> Why should I now hypothesize, in addition to the orderly series of words, another series of mental elements running parallel? That simply duplicates language with something else of the same kind. (1974, p. 152)

Giving *B* a mentalese token, to associate with a sound in Mandarin, and internal operations could suffice for *B* to mean something by the sound *only if the token and operations themselves mean something to B*. In short, by introducing internal representations and operations we are 'only pushing the question [of what generates meaning] one step further back' (1965, p. 36). If *B*'s public manipulation of the Mandarin sound for *o* is insufficient for him to *mean o*, what makes his internal manipulation of a mentalese token suffice for meaning? In Wittgenstein's view, *B* in fact does not mean anything at all; a non-Chinese speaker's utterance of a Chinese sentence is a paradigm example of talking 'without thinking' (1989b, p. 7) and does not even count as a genuine speech act (p. l54; see Proudfoot, 2002). According to Wittgenstein, for an individual's utterance of an expression to have meaning, the individual must have an appropriate history of having learned the expression and must use the expression in other contexts (see Section 5.4). *B* does not satisfy this requirement.

This leaves the alternative that I am *not conscious* of the internal representations and operations in virtue of which I mean and think about objects. However, the problems just described arise for this possibility too. The AP model appears to leave Wittgenstein's question ('What explains intentionality?') unanswered.

3.3. Can meaning be determinate?

On the AP model, what the speaker means is determined by internal processes and representations. But, Wittgenstein asks: how can this thesis avoid the consequence that meaning is *indeterminate* (i.e., that language is ambiguous)?

Wittgenstein pointed out that, for every representation *r*—internal or external, immaterial or physical—and every object *o*, on some interpretation the former stands for the latter. A photograph of Wittgenstein may, if so designated in a poker game, stand for $5; it may, if so designated in a discussion of the layout of Cambridge, stand for the town railway-station; and so on. He gave a now-famous example:

> I see a picture; it represents an old man walking up a steep path leaning on a stick.—How? Might it not have looked just the same if he had been sliding downhill in that position? (1972, p. 54)

What, then, determines the content of any particular internal representation (for example, one in mentalese)? This question arises even in the case of representations that seem to carry their content (as we might say) on their face:

> I have the intention of carrying out a particular task and I make a plan. The plan in my mind is supposed to consist in my seeing myself acting thus and so. But how do I know, that it is myself that I'm seeing? … [W]hy do I call it the picture of *me*? (1974, p. 102)

The traditional answer is that a representation stands for a particular object just in virtue of *resembling* that object. Thus a traditional motive for introducing internal representations as accompaniments to language is the hypothesis that the former but not the latter resemble (and so represent) their objects. This was Locke's view: ideas stand for the objects in the world that they naturally resemble, and words are linked to ideas by convention. Wittgenstein argued that this answer would not do: 'It is not similarity that makes the picture a portrait (it might be a striking resemblance of one person, and yet be a portrait of someone else it resembles less).' (1974, p. 102)

The traditional answer is unsatisfactory just because, for any representation *r* and any object *o*, on some interpretation *r* resembles *o*. For example, the photograph of Wittgenstein resembles not only Wittgenstein but also a photograph of Bertrand Russell (e.g., in the respect that both are photographs of men), a photograph of Trinity College (e.g., in the respect that both are black-and-white photographs), my highly polished desk (in that both are shiny), a shoebox (in that both are rectangular), and so on. Likewise for internal representations: if the internal representation is to determine the object of the mental state by resembling the object, the respect in which representation and object are similar must first be determined. How is this to be done? It cannot be done by means of an *additional* internal representation that determines the similarity space, for that would merely initiate a regress of representations ('[T]he fact that I saw *him* can't have been in the picture that was in my mind's eye. So was there a picture *and* a thought there?' (1990, §808)).

Pinker agrees that words and pictures are ambiguous, and that similarity is 'in the eye of the beholder' (2005, p. 25). However, he uses the ambiguity of words and pictures to argue that we think in mentalese:

> Pictures are ambiguous, but thoughts, virtually by definition, cannot be ambiguous. … If a mental picture is used to represent a thought, it needs to be accompanied by a caption, a set of instructions for how to interpret the picture—what to pay attention to and what to ignore. The captions cannot themselves be pictures, or we would be back where we started. … [T]here's no getting around the need for abstract symbols and propositions that pick out *aspects* of an object for the mind to manipulate. … [I]mages must give way to ideas. (1999, pp. 297–298)

According to Pinker, mentalese is (what philosophers call) an *ideal* language. Spoken languages fail to include necessary information: 'In contrast, the "language of thought" … can leave nothing to the imagination, because it *is* the imagination' (p. 70; see also 2007, p. 150). Spoken languages also contain redundancy: 'articles, prepositions, gender suffixes, and other grammatical boilerplate. These are needed to help get information from one head to another by way of the mouth and ear, a slow channel, but they are not needed inside a single head where information can be transmitted directly by thick bundles of neurons' (1999, p. 70; see also 2007, p. 150).[8]

However, as Wittgenstein argued, nothing intrinsic to a symbol determines its meaning ('[W]e give one interpretation after another; as if each one contented us at least for a moment, until we thought of yet another standing behind it.' (1972, §201)).[9] As formal model theory teaches us, any abstract system of symbols can be interpreted in many different ways. Pinker's mentalese 'captions' and 'instructions' are no exception. As Wittgenstein remarked, '[H]ow can [an internal symbol] give the sentence a *particular* life, if language doesn't do so? How should it be less ambiguous than the language of words?' (1988a, §677).

Pinker appears to endorse a *causal* theory of meaning (1999, pp. 80–81)—roughly speaking, the hypothesis that a representation stands for an object just in virtue of being in a particular causal relation to it. (It is unclear how this fits with his core claim (Section 2) that meanings are mentalese symbols; see Section 3.5.) Proponents of causal theories, it seems, assume that the relations they posit are determinate and so avoid Wittgenstein's challenge. However, considerable work is required to

[8] For Locke too, language is full of 'imperfections' (1690/1959, III, ix–x); in a perfect language each word signifies a (simple, clear, and distinct) idea.

[9] Wittgenstein does not claim that meaning is *in fact* indeterminate (see Section 5, also Proudfoot & Copeland, 2002); his claim is solely that, contra Pinker, there are no 'precise mental representation[s]' that fix meaning (Pinker, 2007, p. 112). Hence, he should not be confused with the 'radical pragmatists' that Pinker attacks (pp. 107–124).

make this approach satisfactory.[10] Everything (that can stand in a causal relation at all) is causally related to very many other things. For every representation *r*, there are very many objects *o*, such that there is some causal relation between *r* and *o*. (For example, the photograph of Wittgenstein is causally related to Wittgenstein, Wittgenstein's mother, the original photographer, the author who uses the photograph in a biography of Wittgenstein, any reader of this biography, and so on.) How, then, does Pinker guarantee that (as he himself says) 'thoughts … cannot be ambiguous'?

According to Pinker, 'we are not open-minded logicians but happily blinkered humans, innately constrained to make only certain kinds of guesses' about what other people mean (1995, pp. 153–154):

> The unavoidable implication is that a sense of "similarity" must be innate. This much is not controversial: it is simple logic. (p. 416)

This conclusion, Pinker claims (p. 153), solves Quine's famous problem of the indeterminacy of translation (namely, how should an anthropologist translate a native's utterance, given different possible and behaviourally indistinguishable interpretations?).

However, Pinker's conclusion fails to solve Wittgenstein's problem (namely, what, if anything, determines the content of an internal representation?). There are two reasons for this. First, Pinker takes Quine's problem to be epistemic and his own solution is also epistemic: it is intended to explain how a language-learner *knows* what a competent speaker means. In contrast, the indeterminacy that concerns Wittgenstein is not merely epistemological (he remarked, 'If God had looked into our minds he would not have been able to see there whom we were speaking of' (1972, p. 217).) As a result, Wittgenstein's question remains: what, if anything, determines what the competent speaker means? Second, Pinker claims that the language-learner is innately constrained to translate a word (heard for the first time) into the *same* internal representation as the one in the competent speaker's head. But this is to say that the representations in the heads of the learner and the competent speaker have the same *content*, and what makes this so? Again, Pinker's version of the AP model is 'only pushing the question [in this case, of what *determines* meaning] one step further back' (Wittgenstein, 1965, p. 36).

3.4. Summing up

The objections in Sections 3.1–3.3 to the AP model's account of meaning justify our looking elsewhere for an analysis of meaning—at accounts, that is, where thought is *not* primary. In case this seems at odds with contemporary psychology, it should be noted that Wittgenstein's arguments target the AP model's claim that representation *consists in* inner processing (and the related supposition that scientific investigation of such processing is investigation of *meaning*). His objections have no implications for the assumption that neural machinery is the *causal* basis of language, to be investigated by scientific methods. In fact, Wittgenstein anticipated recent non-representationalist approaches within cognitive science (see Section 5.3).

The AP model may, nevertheless, still seem intuitive. So, for the sake of argument, let us assume that the problems in 3.1–3.3 have been solved. (This is unlikely; the last problem alone has generated the extensive philosophical literature on 'meaning scepticism'.) Let us suppose that I have an internal representation token that, according to our hypothesized theory of representational content, has meaning. Moreover, it has a fixed meaning; it stands for Wittgenstein (rather than, say, Russell). And I am conscious of this internal representation. (To make the point vivid, let this representation be an image of Wittgenstein.) Is all this sufficient, as the AP model claims, for *me* to mean Wittgenstein?

[10] So too other contemporary theories of representation. For example, on *co-variationist* (tracking) theories, a representation typically represents the object with which it co-varies. However, one of Wittgenstein's remarks on intention (1974, p. 142) anticipates Fodor's (1987) 'disjunction problem' for these theories. According to *teleological* theories, the content of a representation is determined by the evolutionary history of the organism. This is Millikan's (1990) solution to what she calls 'the Kripke–Wittgenstein paradox'. However, Wittgenstein made an important distinction (shared by many contemporary cognitive scientists) between following a rule and merely behaving in accordance with a rule (1972, §232). The latter is insufficient for meaning (see Proudfoot, 2004a, 2004b).

Wittgenstein recommended (as a prophylactic against magic-invoking theories) that we investigate philosophical claims about the mind by replacing, in our analysis, 'any working of the imagination by acts of looking at real objects' (1965, p. 4). It seems clear that, when I am looking at a real portrait of Wittgenstein, I need not be *thinking* about Wittgenstein. Perhaps the portrait of Wittgenstein always makes me think of Russell. I may say to myself, 'He lived an extraordinary life'. It seems clear that in doing so I mean *Russell*, regardless of the fact that I am looking at a portrait of Wittgenstein. But if this is so, then what reason is there to claim that possession of an *internal* portrait of Wittgenstein would suffice for me to mean Wittgenstein, for example when I say 'He lived an extraordinary life'? The AP model needs to answer this question.

If the proponent of the AP model does not answer this question, then she must explain the intentionality of internal physical processes *without* deriving it from the content of embedded internal representations. Yet theories of meaning typically introduce internal representations in order to explain the supposed intentionality of these processes (Section 3.2). On the other hand, let us suppose that, according to the proponent of the AP model, possession of an internal representation (but not looking at a real portrait) *is* sufficient for me to mean Wittgenstein. But now the explanatory work seems to be being done, not by the *representation* of Wittgenstein, but by the fact that this representation is a *mental* one. The AP model appears to explain meaning only by smuggling it into the (unexplicated) 'mind'.

3.5. Pinker's rejection of the AP model

In his latest book (Pinker, 2007), Pinker departs from central elements in the AP model's account of meaning (he does not give much indication that his earlier and current views differ). Now his view in important respects is more like Wittgenstein's: the meanings of words are *not* ideas, and instead depend upon the *external* world (on Wittgenstein's externalism see Section 4; also Proudfoot, 2004a,b). Now for Pinker, 'a word must be more than a shared definition and image' (2007, p. 9); 'words for many kinds of things … are rigidly yoked to the world by acts of pointing, dubbing, and sticking (p. 11). He accepts a version of a Kripkean direct-reference theory: proper names, natural-kind terms, and the names of artefacts are rigid designators, whose meanings (at least in the 'wide' sense) just are the objects in the world at the end of the relevant causal chains (pp. 9–13, 281–296). He also accepts Putnam's famous thesis that '"meanings" just ain't in the head', but depend on the social and physical environment (p. 289). In addition, whereas Pinker in his earlier work talked of language 'convey[ing] a concept from mind to mind' and of its being 'the conduit through which people share their thoughts' (Section 2), now he remarks 'Another misleading conceptual formula is the conduit metaphor, in which to know is to have something and to communicate is to send it in a package' (2007, p. 84).

Pinker has not, however, let go entirely of this part of the AP model. He retains the view that 'language is above all a medium in which we express our thoughts and feelings' and that 'thoughts must be much finer-grained than words' (2007, pp. 24, 150). In some places, he appears still to claim that the meanings of words are mental items. He says, 'the meanings of words and sentences are formulas in an abstract language of thought' (p. 125) and, strikingly, 'language works by evoking meanings, and … meanings are continuous with thoughts' (p. 128). Even in the case of proper names and natural-kind terms, Pinker claims, '*some* part of a word meaning must be in people's heads' (p. 290). One way to understand Pinker's position is to compare him again with Locke. Although for Locke the meanings of words are ideas in the speaker's head, these ideas in turn are to stand for objects in the world. Pinker's position appears similar; he claims that 'the human mind contains representations of the meanings of words' (p. 90). This is an uncomfortable stance which appears to straddle two opposing models of language.

In Pinker's latest book, Wittgenstein is cited solely as a proponent of linguistic determinism. Pinker shows no awareness of Wittgenstein's comprehensive attack on the hypothesis that meanings are ideas or of his emphasis on community and environment—the core elements of Pinker's new stance. This is unfortunate, since Pinker discusses the relation between language and thought at length; paying attention to Wittgenstein would have given him, not only more reason to reject the AP model, but also more options to put in its place.

4. Wittgenstein's account of the relation between language and thought

In this section, I discuss the fundamentals of Wittgenstein's alternative to the AP model's account of meaning. Wittgenstein claimed that (1) a thought has content only in that its 'expression' has meaning (Section 4.1); and (2) for a symbol to have meaning is for it to have a role in a 'language-game' (4.2). (1) and (2) together yield the conclusion: for a thought to have content is for its expression to have a role in a language-game. According to Wittgenstein, it is our linguistic practices that give meaning, and 'a *particular* life', to a symbol and thereby to a thought: 'it is the *system* of language that makes the sentence a thought and makes it a thought *for us*' (1988a, §677; 1974, p. 153). This is his solution to (what he called) 'the whole problem of representation' (1974, p. 102).

Wittgenstein's approach is very different from the AP model (Section 4.3). In Wittgenstein's view, words give meaning to thought; on the AP model, thought gives meaning to words. He said cryptically, 'It is *in language* that it's all done' (1974, p. 143). I shall refer to Wittgenstein's approach as the *intra-linguistic model* of the relation between language and thought (for a detailed account of this model see Proudfoot and Copeland, 2002). Returning to the starting-point of this paper, *does language shape thought?* On the intra-linguistic model, the answer is *yes*: the content of a thought is just the meaning of its 'expression'.

4.1. Thought and its 'expression'

On the AP model, meaning is a mental act or process (i.e., a phenomenon occurring at a specific time and place). Wittgenstein said, in contrast, that meaning is *dispositional*.[11] The speaker who (in writing or uttering the symbol 'N') *means* N—for example, who in uttering 'bread' means bread—is someone who is *following a rule* for the use of 'N'.[12] Rule-following is dispositional. And so the speaker who (in writing or uttering 'N') means N is someone who is disposed to act in certain sorts of ways:

> 'So you meant that man *when you uttered the sentence*.'—Yes, but only in the kind of way that I then knew also that $6 \times 6 = 36$. (1974, p. 103)

Which dispositions are involved here? Paradigmatically, the speaker is disposed to respond to the question 'Whom do you mean?' by indicating N, for example by saying 'N is an old school friend' (just as the arithmetical rule-follower is disposed to respond to the question 'What is 6×6?' by saying '36'). A vocaliser without the appropriate dispositions would simply not be speaking a language, any more than a putative arithmetical rule-follower would know the $6\times$-table.

Wittgenstein's account of rule-following is *externalist*: which rule a speaker is following is determined by her behaviour in the context of history and environment. When the context is clear, the speaker's behaviour is sufficient to determine the content of her utterance, and so too of her thought. The contrast with the AP model is marked: in Wittgenstein's account, behaviour, history, and environment play the role taken in the AP model by internal processes and representations.

But what of the situation where, we wish to say, I am thinking—about N, for example—but I do *not* write or utter any symbols (even as part of an interior monologue)? Perhaps this is an example of what Pinker means by 'pure thought' (2007, p. 124). In Wittgenstein's view, here too I am disposed to act in certain sorts of ways—paradigmatically, to speak:

> Imagine a person who is taking a break in his work, and is staring ahead seemingly pondering something, in a situation in which we would ask ourselves a question, weigh possibilities—would we necessarily say of him that he was reflecting? Is not one of the prerequisites for this that he *be in command* of a language, i.e., be able to express the reflection, if called upon to do so? (1980b, §185)

[11] Wittgenstein acknowledged that '[i]n certain of their applications the words "understand", "mean" refer to a psychological reaction while hearing, reading, uttering etc. a sentence' (1974, p. 41). In addition, although 'what happens' while uttering a word (or looking at a picture) does not determine mental content, 'how one may use' statements about what happens can play a role in doing so (1988a, §315).

[12] This includes uttering 'N' silently, in the course of inner speech.

Thus, Wittgenstein remarked 'The inner is tied up with the outer not only empirically, but also logically' (1993a, p. 63). Even 'pure thought', contra Pinker, is tied up with language.

The place of dispositions in Wittgenstein's account of meaning and thinking has a crucial consequence. Meaning is dispositional, but it is not a disposition to report or express some internal process or representation of which we could predicate content—for Wittgenstein, the arguments in Section 3 have eliminated this possibility. *So what are we to predicate content of?* The only candidate, it appears, is the *behaviour* to which the thinker is disposed:

> When I expect someone,—what happens? I perhaps look at my calendar and see his name against today's date and the note '5 p.m.' I say to someone else 'I can't come to see you today, because I'm expecting N'. I make preparations to receive a guest. I wonder 'Does N smoke?', I remember having seen him smoke and put out cigarettes. Towards 5 p.m. I say to myself 'Now he'll come soon' … This and many other more or less *similar* trains of events are called 'expecting N to come'. (1974, p. 141)

The content of the expectation is the meaning of these 'expressions' (see Section 4.3). Linguistic behaviour is the most salient 'expression' of thought.

4.2. Meaning and 'language-games'

But how does linguistic behaviour gain meaning, if not (as on the AP model) by conveying thought? Wittgenstein's famous answer is that for a symbol to be meaningful is for it to have a role in a *language-game*: 'The sign … gets its significance from the system of signs, from the language to which it belongs' (1965, p. 5).[13] (Wittgenstein used the term 'language-game' for emphasis. In his view, higher-level linguistic performance emerges from the simple interactive behaviour in primitive but complete 'games'—or microworlds, as we might say. In learning language we acquire prototypes and exemplars; the different instances of a concept are linked, not by satisfying a shared definition, but by the 'family resemblances' that we find amongst games.)

Analogously, in Wittgenstein's view, for a symbol to have a particular meaning is for it to have a particular role. This has the consequence that questions of the form 'What do you mean?' are inquiries, not about internal representations, but about the speaker's use of symbols:

> An answer to the question 'How is that meant?' exhibits the relationship between two linguistic expressions. So the question too is a question about that relationship. (1974, p. 45)

This account of meaning provides a solution to the problem of the apparent indeterminacy of representations (Section 3.3). Wittgenstein remarked:

> Our difficulty could be put this way: We think about things, — but how do these things enter into our thoughts? We think about Mr. Smith; but Mr. Smith need not be present. A picture of him won't do; for how are we to know whom it represents? In fact no substitute for him will do. Then how can he himself be an object of our thoughts? (1965, p. 38)

His answer is:

> We want that the wish that Mr Smith should come into this room should wish that just *Mr Smith*, and no substitute, should do the *coming*, and no substitute for that, *into my room*, and no substitute for that. But this is exactly what we said. (p. 37)

This remark may appear cryptic, but on the intra-linguistic model it is a simple truism that 'this is exactly what we said'. For a mental state to be the thought that Mr Smith should come just is for its

[13] This leads to another objection to the AP model: a mysterious mental 'accompaniment' simply cannot provide this system and so 'there is no … point in *postulating* the existence of a peculiar kind of mental act alongside of our expression' (Wittgenstein 1965, p. 42).

'expression' to be the sentence 'Mr Smith should come' or some synonymous symbol. And for this sentence to have the appropriate meaning is just for it to have a certain role in a language-game.

4.3. Comparing the AP model and the intra-linguistic model

On the AP model, the meaning of an utterance is given by the accompanying thought. According to Wittgenstein, as we have just seen, things are almost exactly the other way around:

> The content of the experience just is to be described by the *specific expression* (of the experience). (1980a, §105)

To use Wittgenstein's own example (Section 3.2), how can some hypothesized internal act amount to an act of *wishing*? For Wittgenstein, it is the 'expression' of this supposed act, typically linguistic behaviour, that has meaning (i.e., a role in a language-game):

> [I]f I imagine the expression of a wish as the act of wishing, the problem [of intentionality] appears solved, because the system of language seems to provide me with a medium in which the proposition is no longer dead. (1974, p. 149)

In the same way, a wish is a particular wish in virtue of its expression, and not, as on the AP model, in virtue of some internal representation or operation in the thinker:

> Where are we to find what makes the wish *this* wish, even though it's only a wish? Nowhere but in the expressed wish. (p. 150)

However, talk of 'expressions' of thought may seem simply to reintroduce the AP model. On this model, language expresses or (to use Locke's and Pinker's word) *conveys* thought: 'The most remarkable aspect of language is its *expressive* power: its ability to convey an unlimited number of ideas from one person to another via a structured stream of sound.' (Pinker, 2004, pp. 139–140). In Wittgenstein's view, the AP model completely misconstrues the notion of the expression of thought:

> Misleading parallel: the expression of pain is a cry—the expression of thought, a proposition. As if the purpose of the proposition were to convey to one person how it is with another: only, so to speak, in his thinking part and not in his stomach. (1972, §317)

According to Wittgenstein, the AP model's hypothesis '[t]hat *pure thought* is conveyed by words and is something different from the words' is 'a *superstition*' (1979, p. 54). Moreover, this superstition leads to an inflated ontology: '"thinking" is … [an] imaginary auxiliary activity' (Wittgenstein, 1988b, §226).

For Wittgenstein, a thought *is* 'expressed' in an utterance, but not in the AP model's sense that the utterance conveys the content of some independent internal state in the speaker. It is expressed only in the sense that the utterance (or gesture) is *criterial* for the attribution of the thought to the speaker—all things being equal, we assign that thought to the speaker when he or she utters those words (or makes that movement). In his view, for example:

> Knowledge is not translated into words when it is expressed. The words are not a translation of something else that was there before they were. (1970, §191)

(In the remainder of my discussion of Wittgenstein, 'expression' should be read in this way.)

On the AP model, language is merely a *guide* to cognition. For Pinker, language is 'a window into human thought, which is couched in a richer and more abstract format' (2007, pp. 91–92). On the intra-linguistic model, in contrast, language *displays* thought—just as facial 'expressions' do (Wittgenstein said, 'Look into someone else's face and see the consciousness in it … You see on it, in it, joy, indifference, interest, excitement, dullness etc.' (1988a, §927)). For Wittgenstein, 'there is no more direct way of reading thought than through language': 'Thought is not something hidden; it lies open to us.' (1980a, p. 26).

In the next section, I turn to Pinker's criticisms of Wittgenstein and of (what Pinker calls 'radical' versions of) the hypothesis that language shapes thought.

5. Pinker, Wittgenstein, and the prisonhouse hypothesis

Pinker includes Wittgenstein amongst proponents of the hypothesis that language is a 'prisonhouse of thought' (2003, p. 210; for short I refer to this as the 'prisonhouse hypothesis'):

> Rather than being appreciated for its ability to communicate thought, [language] was condemned for its power to *constrain* thought. … "The limits of my language mean the limits of my world," wrote Ludwig Wittgenstein. (p. 207)
> [P]eople simply assume that words determine thoughts. … Philosophers argue that since animals lack language, they must also lack consciousness—Wittgenstein wrote, "A dog could not have the thought 'perhaps it will rain tomorrow'"—and therefore they do not possess the rights of conscious beings. (1995, p. 56)

In this section, I argue that, first, in these comments Pinker simply misunderstands Wittgenstein, and, second, Pinker's objections to the prisonhouse hypothesis are unsuccessful against Wittgenstein's model of the relation between language and thought.

5.1. Pinker's arguments against the prisonhouse hypothesis

Pinker does not give a precise statement of his hypothesis that language is a prisonhouse of thought (his targets include the 'Standard Social Science Model', 'linguistic determinism', and 'linguistic relativity'). Broadly speaking, he rejects the idea that thought *depends* on, or is *constrained* by, language. He opposes three different theses. (1) *Thought is just (conditioned) linguistic behaviour.* This is behaviourism, the thesis that 'the contents of cognition—ideas, thoughts, plans, and so on—[are] really phenomena of language, overt behavior that anyone could hear and write down' (2003, p. 24). This behaviour also includes words 'muttered silently' (2007, p. 124). (2) *Language is the 'medium' of thought.* This is the thesis that 'people literally think in English, Cherokee, Kivunjo' (1995, p. 56; see also 2007, p. 133). (3) *Language 'determines' thought.* As a (radical) thesis, this claims that 'people's thoughts are determined by the categories made available by their language' (1995, p. 57)—that 'the language we speak makes it difficult or impossible to think certain thoughts, or alters the way we think in surprising or consequential ways' (2007, p. 125). (This claim is the core of Pinker's several versions of 'genuine Linguistic Determinism' (2007, pp. 133–134).) In Section 5.3, I argue that, despite any superficial similarities, Wittgenstein's intra-linguistic model is very different from each of theses (1)–(3).

Of Pinker's various arguments against the prisonhouse hypothesis (i.e., theses (1)–(3), and linguistic determinism in particular), the most promising as difficulties for the intra-linguistic model are the following[14]:

(a) *Observational evidence of thinking without language*. Human examples include: 'babies, who cannot think in words because they have not learned any'; 'deaf children'; 'deaf adults, occasionally discovered, who lack any form of language whatsoever'; 'the fully intelligent aphasic'; 'human adults who, whether or not they think in words, claim their best thinking is done without them' (1995, pp. 67–68). Non-human animal examples include 'monkeys, who cannot think in words because they are incapable of learning them' (1995, p. 68)—here Pinker cites recent studies of object individuation in rhesus macaques (2007, p. 137). (By 'words' Pinker intends natural human languages: non-human communication systems are very differently designed, he claims (in e.g., Pinker, 1997, pp. 182–184).)

(b) *Phenomenological descriptions* of 'a constant give-and-take between the thoughts we try to convey and the means our language offers to convey them' (2003, p. 210). For example:

> [T]he experience of uttering or writing a sentence, then stopping and realizing that it wasn't exactly what we meant to say. To have that feeling, there has to be a "what we meant to say" that is different from what we said. (1995, p. 57)
> Sometimes it is not easy to find *any* words that properly convey a thought. (1995, p. 58)

[14] I discuss Pinker's criticisms of recent attempts to base linguistic determinism on empirical results in 'Empirical Evidence for Neo-Whorfianism? A Wittgensteinian Perspective' (in preparation).

(c) *The theoretical necessity of a distinction between a sign and its meaning*. For example:

> When we hear or read, we usually *remember the gist*, not the exact words, so there has to be such a thing as a gist that is not the same as a bunch of words. (1995, p. 58, emphasis added; see also 2007, p. 149)
> How [else] could *translation* from one language to another be possible? (1995, p. 58, emphasis added)
> The very existence of *ambiguous* sentences, in which one string of words expresses two thoughts, proves that thoughts are not the same as strings of words. (2003, p. 211, emphasis added)

Pinker also mentions the 'ineffable thoughts' that, he claims, a listener may have to use in order to make sense of a new expression (e.g., 2003, p. 209).

(d) Last, Pinker also suggests that *the development of language*, in both the species and the individual, is impossible on the prisonhouse hypothesis:

> If people had trouble thinking without language, where would their language have come from—a committee of Martians? (2007, p. 149)
> How could a child learn a word to begin with? (1995, p. 58; see also 2007, pp. 148–9)

Pinker appears to treat the evidence listed here as conclusive.

5.2. Pinker on Wittgenstein

Pinker frequently uses his first quotation from Wittgenstein ('The limits of my language …') to illustrate the prisonhouse hypothesis. This quotation is from the *Tractatus* (Wittgenstein, 1961). Wittgenstein's account of the relation between language and thought in this early work is significantly different from the intra-linguistic model described in Section 4. According to his early account (the 'picture theory'), language (sentences) and thought (propositions) are 'logical pictures' of reality: they share the 'logical form' of the facts they represent. Contrary to Pinker's claim, in the *Tractatus* language does not 'constrain' thought. In fact, a few sentences after the remark Pinker quotes, Wittgenstein said:

> We cannot think what we cannot think; so what we cannot think we cannot *say* either. (1961, p. 57)

In quoting Wittgenstein, Pinker omits the emphasis in the original, in doing so changing Wittgenstein's meaning. Wittgenstein actually said:

> *The limits of my language* mean the limits of my world. (1961, p. 56)

What Wittgenstein meant is that it is the *world* that limits language (by determining which sentences have sense). Pinker takes Wittgenstein's remark almost exactly the other way around. He classes Wittgenstein with 'the radical versions of the Whorfian hypothesis' (2007, p. 133) and claims that for Wittgenstein language exerts a 'stranglehold' on thought (2003, p. 207). But, in the remark Pinker quotes, all Wittgenstein is claiming is that a speaker cannot say anything nonsensical (i.e., 'beyond the limits of the world' (1961, p. 57)). This is not much of a constraint on thought; in fact, it is almost a tautology.

Pinker also misunderstands his other quotation from Wittgenstein ('A dog could not have the thought …'). Wittgenstein said, 'There is nothing that astounding about a certain concept only being applicable to a being that, e.g., possesses a language' (1988b, §310). He meant that we make propositional-attitude attributions to non-linguistic animals that are inappropriate, since the attitudes concerned require conceptual resources unavailable to the animal—for example, 'A dog can expect its master, but can it expect its master will come the day after tomorrow?' (Wittgenstein, 1990, §360). In fact, Pinker makes almost exactly the same point:

> [T]here are domains of human concepts which are probably unlearnable without language … For example, the notion of a "week" depends on counting time periods that cannot all be perceived at once; we doubt that such a concept could be developed or learned without the mediation of language. (Pinker and Jackendoff, 2005, p. 206)

Moreover, Wittgenstein did not assert that non-human animals lack consciousness.[15]

In sum, Pinker's quotations give no reason to think that Wittgenstein endorsed the prisonhouse hypothesis. In fact he did not, as I argue next.

5.3. Is Wittgenstein's intra-linguistic model a version of the prisonhouse hypothesis?

In attacking this hypothesis, Pinker's first target is *behaviourism* (Section 5.1). According to Pinker, 'In behaviorism, … there was *no such thing* as a talent or an ability. Watson had banned them from psychology, together with other contents of the mind, such as ideas, beliefs, desires, and feelings' (2003, p. 19). Wittgenstein has often been described as a behaviourist and it is true that he said 'The "inner" is a delusion' (1993a, p. 84). Yet he repeatedly denied that he was a behaviourist. For example, he famously said, of pain-behaviour with and without pain, 'What greater difference could there be?' (1972, §304).

The intra-linguistic model is not behaviourist (at least in the standard philosophical or psychological senses). To claim that the content of thought is the meaning of its expression is not to define thought as dispositions to speech, nor to explain and predict linguistic behaviour by means of conditioning. In fact Wittgenstein's philosophical psychology anticipated more recent research into the mind (see Proudfoot, 2004b). He prepared the philosophical ground for connectionist (see Stern, 1991) and dynamicist approaches to cognition. For example, he hypothesized that the 'jottings' in our nervous system that may in fact be required for recall of a text are 'not connected by rules with the words of the text'; they are 'not … a translation with another symbolism' and the 'text [is] not … *stored up* in the jottings' (1970, §612). He said that it is 'impossible to read off thought-processes from brain-processes'; 'this can only be done from the *history* of' our talking and writing (§608)—just as we now say that the functions computed by a neural network cannot be gleaned from the patterns of connectivity it displays, in the absence of knowledge of the training inputs. Wittgenstein was also an early proponent of what is now known as 'situated cognition' (e.g., Clark, 1997) and his theoretical psychology resembles that of the advocates of 'nouvelle AI' such as Brooks (1999). His famous work on family resemblance, having influenced theories of categorization (Rosch, 1987) and neural architectures (Smolensky, 1991), is now of wider significance within AI for the investigation of machine learning and 'knowledge-representation' (Rissland, 2006).

Wittgenstein's remark that '[t]he "inner" is a delusion' is what he called a 'signpost', rather than a statement of theory. His aim was to 'erect signposts at all the junctions where there are wrong turnings so as to help people past the danger points' (Wittgenstein, 1980b, p. 18)—those places in a philosophical investigation where inherited, unnoticed assumptions may skew inquiry. This particular signpost is intended to divert us from unhelpful internal-state analyses of cognition, such as the AP model's failed explanation of meaning. Nevertheless, given Wittgenstein's arguments, Pinker's 'contents of the mind' are indeed unreal. For Pinker, these are internal mental states that can be individuated independently of their expressions (and, in his early work at least, that determine the meaning of words); as such, for Wittgenstein they are 'grammatical fiction[s]' (1972, §307)—metaphysical misinterpretations of our everyday talk of 'states of mind' and 'mental processes' (see below). A scientific psychology need not disagree, and in fact might welcome this demythologizing of the mental.

Pinker's second target is the thesis that language is the *medium* of thought—that is, cognitive processes are operations upon natural-language expressions. It should be evident that Wittgenstein does not endorse this thesis, since he rejects any version of the AP model.

[15] Wittgenstein's position on consciousness, and who has it, is complex. For example, he said:

> [I]f we talk of 'consciousness' or 'I am conscious' as marks of 'conscious being', then we have a misuse of the subject, and 'consciousness' seems something gaseous. (1993b, p. 367)

Contra Pinker, Wittgenstein does not say that animals 'lack consciousness'; saying so would involve precisely this misuse of 'consciousness'.

Pinker's third target is the thesis that language *determines* thought. However, the only sense in which, on the intra-linguistic model, language can be said to 'determine' thought is this: the content of a thought is the meaning of its expression (Section 4). This is a very different claim from the one Pinker attacks.

The differences can be summarized as follows. According to Pinker, in the 'radical' versions of linguistic determinism 'the direction of causation runs from language to thought'; differences in thinking between speakers of two different languages 'must be *caused* by the language' (2007, pp. 133, 136). However, in Wittgenstein's account the relation between language and thought is not causal, but *logical*: thoughts are individuated by their expressions. According to Pinker, too, in 'genuine' linguistic determinism 'the ineffability of a concept in one's mother tongue creates a permanent blind spot in one's ability to entertain it' (p. 133). Wittgenstein is not committed to this: in his account, one's mother tongue is merely one sort of 'language'—that is, one sort of 'expressive' behaviour. And last, according to Pinker, in linguistic determinism 'typological differences alone … [can] channel the thoughts of the respective societies and speakers in different directions' (p. 140). In the only sense in which, for Wittgenstein, language determines thought, 'languages' are not to be differentiated in this way; different languages use different 'tools' (1972, §23) and consist of different language-games.[16]

In the next section I consider whether Pinker's arguments against the prisonhouse hypothesis (Section 5.1) undermine Wittgenstein's much more subtle version of the claim that language determines thought. I begin by examining some of Wittgenstein's remarks concerning the development of language and thought.

5.4. Do Pinker's arguments provide evidence against Wittgenstein's intra-linguistic model?

On the intra-linguistic model, thought *cannot* develop before language, since the content of a thought is just the meaning of its expression. This result is entirely independent of answers to such questions as: How is natural language to be defined, if at all? Is language unique to the human species? Is it an adaptation or a spandrel? It might seem that this issue is an empirical one. But not according to Wittgenstein; it is a simple consequence of the intra-linguistic model, and empirical evidence is irrelevant. How plausible is this? Surely, someone might say, in theorizing about language empirical evidence is everything. Notably, not according to Pinker either, who argues for innate similarity constraints on the grounds of 'simple logic' (Section 3.3).

There are some similarities between Wittgenstein's and Pinker's views of the development of language. Wittgenstein's notion of a human 'form of life' resembles the notion, used by Pinker, of the cognitive niche in which language evolved. For Wittgenstein, as for Pinker, language involves instinct.[17] However, Wittgenstein and Pinker also differ greatly. According to Wittgenstein, it seems, language begins with gestures rather than ideas; indeed, he can be regarded as offering an a priori argument for the gestural origin hypothesis of language.[18] In Wittgenstein's view, what is needed for language to develop is certain sorts of shared (instinctual) behaviour (1988b, §205): 'Our language-game is an extension of primitive behaviour.' (1970, §545). He claims, contra Pinker, that the primitive behaviour from which natural language develops is itself 'the prototype of a way of thinking and not the result of thought' (§541): '[A] language-game does not have its origin in *consideration*. Consideration is part of a language-game.' (§391). He makes a similar claim in the case of the individual language-learner: 'The child learns first to express a wish, and only later to make the supposition that it wished for such-and-such.' (1988a, §478).

[16] This way of differentiating languages is not to use 'the word *language* in a loose way to refer to meanings'—Pinker's criticism of one version of linguistic determinism (2007, p. 128). As Pinker points out, using 'language' in this way would make the thesis that language 'determines' thought vacuous.

[17] Pinker objects to the Standard Social Science Model's emphasis on the *malleability* of human nature (1995, p. 406). The sense in which, for Wittgenstein, language determines thought has no such emphasis. In fact, Wittgenstein said that language-games depend on very general facts about human nature.

[18] Although a priori, Wittgenstein's argument is consistent with studies showing links between the neural mechanisms of speech and of gesture. See, e.g., Willems & Hagoort (2007).

What distinguishes proto-thought ('the prototype of a way of thinking') from the sorts of intentional states expressed in the complex, sophisticated utterances of typical adult humans (we might call the latter 'conceptual thought')? Wittgenstein said that in the former case:

> we wouldn't have the complete employment of 'think'. The word would have reference to a mode of behaviour. Not until it finds its particular use in the first person does it acquire the meaning of mental activity. (1988b, §230)

I now return to Pinker's objections to the prisonhouse hypothesis (in the order they were presented in Section 5.1) and argue that they leave Wittgenstein's version of the claim that language 'determines' thought untouched.

Observational evidence of thinking without language. It may appear that Wittgenstein denies the possibility of thinking without language: after all, he said '[W]here no language is used, why should one speak of "thinking"?' (1970, §109).[19] However, it is probable that here he meant conceptual thought. The rhesus macaques' ability to individuate objects by simple kinds (and other similar examples of animal cognition) is, in contrast, surely an example only of proto-thought. In addition, it is clear from the above that Wittgenstein interprets 'language' widely: the 'primitive behaviour' that is criterial for 'primitive thinking' can, we may assume, be described as primitive *language*. (This would be the developmental beginnings of language, rather than primitive versions of natural languages.) According to Wittgenstein, animals sometimes classed by psycholinguists as non-linguistic, such as chimpanzees, or as pre-linguistic, such as human infants, exhibit this sort of behaviour. It is possible that they exhibit even more sophisticated 'expressive' behaviour:

> if their trial-making were to take on the form of producing a kind of model (or even a drawing) then … [t]o be sure one could … speak here of an operation with signs. (1988b, §188)

And so for Wittgenstein, contra Pinker, human children and monkeys are *not* language-less creatures. This allows Wittgenstein to agree that they think:

> if we were to see creatures at work whose *rhythm* of work, play of expression etc. was like our own, but for their not *speaking*, perhaps in that case we should say that they thought, considered, made decisions. (§186)
>
> So one might distinguish between two chimpanzees with respect to the way in which they work, and say of the one that he is thinking and of the other that he is not. (§229)

As to Pinker's 'fully intelligent aphasic', that the aphasic is judged intelligent suggests that she too has available forms of expression and so is not language-less in Wittgenstein's sense. (In any case, where the impairment is acquired, the aphasic may have the normal speech dispositions.)

Pinker also mentions (as another example of thought disassociated from language) grammatically fluent speakers with cognitive retardation syndromes, including Williams syndrome (2004, p. 145). There is in fact debate concerning the extent of language proficiency in Williams syndrome (WS).[20] Nevertheless, for the sake of argument, suppose Pinker were correct: in WS or in some other developmental disorder, 'substantially retarded children may speak fluently and grammatically and do well on tests of grammatical comprehension and judgment' (2004, p. 145). Such an example might seem to present a difficulty for the intra-linguistic model: these speakers are uttering sentences that are criterial for complex thoughts, but intuitively it is incorrect to ascribe those thoughts to these speakers. How can we allow this intuition without reintroducing the AP model of the relation between language and thought? (This is distinct from the much-discussed question whether WS provides evidence of Pinker's hypothesized innate language 'faculty'.)

[19] For example, Carruthers assumes (incorrectly) that Wittgenstein is 'committed to denying that any non-human animals can entertain genuine thoughts' (2002, p. 661).

[20] For example, Thomas & Karmiloff-Smith (2005, p. 70) say that 'language [in Williams syndrome] can be made to look impressive given IQ but is unremarkable given mental age'. I am grateful to Ulrich Müller for alerting me to this debate and relevant literature.

The answer is that Wittgenstein did not take fluency at face value. He gave the example of (hypothetical) individuals who

> can produce perfect proofs of the most complicated theorems in higher mathematics … [but] are so unintelligent that they cannot make the simplest practical calculations. They can't figure out if one plum costs so-and-so, how much do six plums cost, or what change you should get from a shilling for a twopenny bar of chocolate.—Would you say that they had learnt mathematics or not? (Wittgenstein, 1989a, p. 36)

According to Wittgenstein, for genuine calculation the mathematical symbols must be used in 'mufti' (1983, p. 257)—that is, in appropriate ways in the real world. He made the analogous demand with respect to genuine language-use and words. Do Pinker's (hypothesized) grammatically fluent speakers pass this test? If not, they do not present a problem for the intra-linguistic model. (According to some recent research, WS children have a reduced capacity to categorize objects by names, in comparison to typically-developing children and despite their large vocabularies (Nazzi, Gopnik, & Karmiloff-Smith, 2005). This may suggest that WS children fail Wittgenstein's test.)

Phenomenological descriptions. According to Wittgenstein, phenomenological descriptions are fine in their place. *Everyday* talk of thoughts as 'in' us is innocuous (1972, §423). In ordinary contexts (where we are not theorizing) we can treat assertions, wishes, and so on as telling us something about (what we call) our 'state of mind' (Wittgenstein, 1988a, §493–496); we can distinguish words 'as signs of mental processes' from words 'learned by heart', even though 'the expression "as signs of mental processes" is misleading' (§292)—misleading, that is, if we give it a realist interpretation. The intra-linguistic model is compatible with these various phenomenological descriptions.

Pinker's mistake is to think that these everyday descriptions commit us to the AP model. For example, contrary to Pinker, it does not follow from 'realizing that it wasn't exactly what we meant to say' that there is 'a "what we meant to say"'—any more than it follows from my realizing that an apple wasn't what I wanted to eat that there is a determinate something that I do want to eat.

Sign versus meaning and *the development of language.* Wittgenstein's emphasis on the primacy of language does not prevent him from making a distinction between a sign (i.e., a word or other 'expressive' behaviour) and its meaning: the latter, famously, is the *role* of the sign in the language-game. This addresses Pinker's examples (Section 5.1). For instance, I remember the meaning of a text when the words I do (or would) utter have the same role as, even if they are not identical to, the exact 'bunch of words' in the text. And, as we have seen, Wittgenstein's intra-linguistic model also allows for the development of language (the language of even a primitive language-game is rule-governed).

There are problems for use-based theories of meaning, and Wittgenstein's remarks on the development of language are programmatic. Nevertheless, Pinker's own approach is in difficulty here. Wittgenstein's criticisms of the AP model undermine Pinker's claim to have provided an account of meaning, and the 'private language argument' is a direct attack on the Lockean model of the development of language. With respect to the latter, Pinker's scientific sophistication is irrelevant. For example, his 'ineffable thoughts' (Section 5.1) look very like one of Wittgenstein's well-known targets—the 'something about which nothing could be said' (1972, §304). Wittgenstein described an ineffable thought or feeling as 'the private object, the sense-datum, the "object" that I grasp immediately with the mental eye, ear, etc' (1988a, §109). He aimed to demonstrate that language could not develop from such thoughts: in his view, there is simply no theoretical or practical difference between 'a something about which nothing could be said' and nothing at all (1972, §304). At the least, Pinker should address this objection.

6. Conclusion

According to Pinker:

> There are many ways in which language connects to thought, some banal, some radical. … [T]he radical versions are the ones that excite people but the boring versions are the ones consistent with the discoveries (2007, p. 126)

On Pinker's version of the connection between language and thought, language does not shape thought in any philosophically interesting sense. In this paper I have argued that Pinker's account is remarkably Lockean, despite his opposition to Locke. In addition, it is vulnerable to Wittgenstein's powerful criticisms of the 'accompanying picture'. For Wittgenstein, language shapes thought, in a fundamental sense. Wittgenstein's account is very different from Pinker's characterisation of it. It is consistent with the 'discoveries' Pinker cites, and easily survives Pinker's other criticisms of 'radical' approaches to the relation between language and thought.[21]

References

Bermúdez, J. L. (2003). *Thinking without words*. Oxford: Oxford University Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227–247.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*, 481–499.

Brooks, R. A. (1999). *Cambrian intelligence: the early history of the new AI*. Cambridge, MA: MIT Press.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25*, 657–726.

Chadha, M. (2007). No speech, never mind!. *Philosophical Psychology, 20*, 641–657.

Clark, A. (1997). *Being there*. Cambridge, MA: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: the problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Locke, J. (1959). *An essay concerning human understanding*. (Collated and annotated by A.C. Fraser). Dover: New York. (Original work published 1690).

Millikan, R. G. (1990). Truth rules, hoverflies, and the Kripke–Wittgenstein paradox. *The Philosophical Review, 99*, 323–354.

Nazzi, T., Gopnik, A., & Karmiloff-Smith, A. (2005). Asynchrony in the cognitive and lexical development of young children with Williams syndrome. *Journal of Child Language, 32*, 427–438.

Pinker, S. (1995). *The language instinct: how the mind creates language*. New York: HarperPerennial.

Pinker, S. (1997). Evolutionary biology and the evolution of language. In M. Gopnik (Ed.), *The inheritance and innateness of grammars* (pp. 181–208). New York: Oxford University Press.

Pinker, S. (1999). *How the mind works*. London: Penguin.

Pinker, S. (2003). *The blank slate: the modern denial of human nature*. London: Penguin.

Pinker, S. (2004). Language as an adaptation to the cognitive niche. In D. T. Kenrick, & C. L. Luce (Eds.), *The functional mind: readings in evolutionary psychology* (pp. 139–156). Boston: Pearson.

Pinker, S. (2005). Evolutionary psychology and the blank slate. In J. Stangroom (Ed.), *What scientists think* (pp. 23–36). New York: Routledge.

Pinker, S. (2007). *The stuff of thought: language as a window into human nature*. New York: Viking.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95*, 201–236.

Proudfoot, D. (2002). Wittgenstein's anticipation of the Chinese room. In J. Preston, & M. Bishop (Eds.), *Views into the Chinese room: new essays on Searle and artificial intelligence* (pp. 167–180). Oxford: Oxford University Press.

Proudfoot, D. (2004a). Robots and rule-following. In C. Teuscher (Ed.), *Alan Turing: life and legacy of a great thinker* (pp. 359–379). Berlin: Springer-Verlag.

Proudfoot, D. (2004b). The implications of an externalist theory of rule-following behaviour for robot cognition. *Minds and Machines, 14*, 283–308.

Proudfoot, D., & Copeland, B. J. (2002). Wittgenstein's deflationary account of reference. *Language and Communication, 22*, 331–351.

Rissland, E. L. (2006). AI and similarity. *IEEE Intelligent Systems, 21*, 39–49.

Rosch, E. (1987). Wittgenstein and categorization research in cognitive psychology. In M. Chapman, & R. A. Dixon (Eds.), *Meaning and the growth of understanding: Wittgenstein's significance for developmental psychology* (pp. 151–166). Berlin: Springer-Verlag.

Smolensky, P. (1991). Connectionism, constituency, and the language of thought. In B. Loewer, & G. Rey (Eds.), *Meaning in mind: Fodor and his critics* (pp. 201–227). Oxford: Blackwell.

Stern, D. G. (1991). Models of memory: Wittgenstein and cognitive science. *Philosophical Psychology, 4*, 203–218.

Thomas, M., & Karmiloff-Smith, A. (2005). Can developmental disorders reveal the component parts of the human language faculty? *Language Learning and Development, 1*, 65–92.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain and Language, 101*, 278–289.

Wittgenstein, L. (1961). *Tractatus logico-philosophicus*. (D. F. Pears & B. F. McGuinness, Trans.). London: Routledge & Kegan Paul.

Wittgenstein, L. (1965). *The blue and brown books*. New York: Harper.

Wittgenstein, L. (1970). (G. E. M. Anscombe, Trans.). In G. E. M. Anscombe, & G. H. von Wright (Eds.), *Zettel*. Berkeley and Los Angeles: University of California Press.

Wittgenstein, L. (1972). (G. E. M. Anscombe, Trans.). In G. E. M. Anscombe, & R. Rhees (Eds.), *Philosophical investigations* (2nd ed.). Oxford: Basil Blackwell.

Wittgenstein, L. (1974). (A. J. P. Kenny, Trans.). In R. Rhees (Ed.), *Philosophical grammar*. Oxford: Basil Blackwell.

Wittgenstein, L. (1979). In A. Ambrose (Ed.), *Wittgenstein's lectures, Cambridge 1932–1935: from the notes of Alice Ambrose & Margaret Macdonald*. Oxford: Basil Blackwell.

[21] I am grateful to Ulrich Müller, Tim Racine, Jack Copeland, and an anonymous reviewer for their valuable comments on the manuscript of this article.

Wittgenstein, L. (1980a). In D. Lee (Ed.), *Wittgenstein's lectures, Cambridge 1930–1932: from the notes of John King & Desmond Lee*. Oxford: Basil Blackwell.
Wittgenstein, L. (1980b). (P. Winch, Trans.). In G. H. von Wright (Ed.), *Culture and value*. Oxford: Basil Blackwell.
Wittgenstein, L. (1983). (G. E. M. Anscombe, Trans.). In G. H. von Wright, R. Rhees, & G. E. M. Anscombe (Eds.), *Remarks on the foundations of mathematics* (revised ed.). Cambridge, MA: MIT Press.
Wittgenstein, L. (1988a). (G. E. M. Anscombe, Trans.). In G. E. M. Anscombe, & G. H. von Wright (Eds.), *Remarks on the philosophy of psychology, Vol. 1*. Chicago: University of Chicago Press.
Wittgenstein, L. (1988b). (C. G. Luckhardt & M. A. E. Aue, Trans.). In G. H. von Wright, & H. Nyman (Eds.), *Remarks on the philosophy of psychology, Vol. 2*. Chicago: University of Chicago Press.
Wittgenstein, L. (1989a). In C. Diamond (Ed.), *Wittgenstein's lectures on the foundations of mathematics, Cambridge 1939: from the notes of R. G. Bosanquet, Norman Malcolm, Rush Rhees, & Yorick Smythies*. Chicago: University of Chicago Press.
Wittgenstein, L. (1989b). In P. T. Geach (Ed.), *Wittgenstein's lectures on philosophical psychology 1946–47: notes by P. T. Geach, K. J. Shah, & A. C. Jackson*. Chicago: University of Chicago Press.
Wittgenstein, L. (1990). (C. G. Luckhardt & M. A. E. Aue, Trans.). In G. H. von Wright, & H. Nyman (Eds.), *Last writings on the philosophy of psychology, Vol. I*. Chicago: University of Chicago Press.
Wittgenstein, L. (1993a). (C. G. Luckhardt & M. A. E. Aue, Trans.). In G. H. von Wright, & H. Nyman (Eds.), *Last writings on the philosophy of psychology, Vol. II*. Oxford: Blackwell.
Wittgenstein, L. (1993b). In J. Klagge, & A. Nordmann (Eds.), *Philosophical occasions 1912–1951*. Indianapolis: Hackett.