5

Knowing That I Am Thinking

Alex Byrne

> *Soc.* . . . I speak of what I scarcely understand; but the soul when thinking appears to me to be just talking—asking questions of herself and answering them, affirming and denying. And when she has arrived at a decision, either gradually or by a sudden impulse, and has at last agreed, and does not doubt, this is called her opinion. I say, then, that to form an opinion is to speak, and opinion is a word spoken,—I mean, to oneself and in silence, not aloud or to another: What think you?
> *Theaet.* I agree.
>
> Plato, *Theaetetus*

I. Introduction

We often know that we are thinking, and what we are thinking about. Here 'thinking' is not supposed to be an umbrella term for cognition in general, but should be taken in roughly the sense of 'a penny for your thoughts': mental activities like pondering, ruminating, wondering, musing, and daydreaming all count as thinking. In the intended sense of 'thinking', thinking is not just propositional: in addition to thinking *that p*, there is thinking *of* (or *about*) *x*. Belief is necessary but not sufficient for thinking that *p*: thinking that *p* entails believing that *p*, but not conversely.[1]

Consider Kylie:

On summer afternoons in Canberra, the baking sun reflects off Lake Burley Griffin, and the water shimmers. Up behind the university, in the botanical gardens, a cascading stream of water helps to maintain the humidity of the rainforest gully. These are just a couple of Kylie's thoughts on the subject of water, her water thoughts. Amongst Kylie's many other thoughts that involve the concept of water are these: that there is water in the lake, that trees die without water, that

Thanks to audiences at Umeå University, the University of Melbourne, Monash University, UMass Amherst, the Institute of Philosophy (London), and a reading group at the RSSS, ANU. I am especially indebted to Martin Davies for discussion.

[1] In the simple present and past tenses (e.g.), 'think that *p*' is near-enough synonymous with 'believe that *p*', as in 'I think/thought that the pub is/was open'.

water is a liquid and, of course, that water is wet. When Kylie thinks consciously, in a way that occupies her attention, she is able to know what it is that she is thinking. This is true for thoughts about water, as for any other thoughts. (Davies 2000: 384–5)

As the watery clue suggests, Davies uses Kylie to discuss McKinsey's puzzle, the alleged incompatibility of externalism and the apparent fact that we can obtain self-knowledge from the armchair. But that is not this paper's topic. Suppose—if only for the sake of the argument—that McKinsey's puzzle can be solved, and that externalism and armchair self-knowledge are compatible. *How* does Kylie know that she is thinking about water "from the armchair"? Just by sitting comfortably and "introspecting", Kylie can discover that she is thinking *about* water, and that she is thinking *that* trees die without water. But what *is* "introspecting", exactly? And why is it—as is commonly assumed—a particularly effective way of finding out what one is thinking?

II. Privileged and peculiar access

In his classic paper "Anti-Individualism and Privileged Access", McKinsey starts by saying that

> [i]t has been a philosophical commonplace, at least since Descartes, to hold that each of us can know the existence and content of his own mental states *in a privileged way* that is *available to no one else*. (McKinsey 1991: 9, emphasis added)

This "philosophical commonplace" consists of two distinct claims, which McKinsey does not clearly distinguish. The first is that we have *privileged* access to the existence and content of our mental states. Approximately put: beliefs about one's mental states acquired through the usual route are more likely to amount to knowledge than beliefs about others' mental states. Kylie's belief that she is thinking about water is more likely to amount to knowledge than our belief that she is thinking about water. We might judge that Kylie is thinking about water because we see her gazing pensively at Lake Burley Griffin, or because we overhear someone we take to be Kylie muttering "Water is a liquid". We can easily go wrong in an ordinary situation of this kind. Perhaps when gazing at the lake Kylie is not thinking of water, but of the original Canberra Planner, Walter Burley Griffin himself; or perhaps Kylie is not the person muttering.

Kylie herself cannot go wrong as easily, or so it is natural to suppose. If we offer Kylie a penny for her thoughts, and she sincerely replies "I was thinking about water", it would be bizarre for us to doubt Kylie's claim. And this is not because we have a blanket policy of taking avowals of mental states to be beyond dispute—we plainly don't. Setting factive and object-entailing states (e.g. *regretting drinking five pints*, and *seeing an echidna*) to one side, we make subtle distinctions between our epistemic access to mental states in ordinary life. It would not be at all unexceptional, for instance, for us to wonder whether Kylie is mistaken about what she wants. Perhaps Kylie falsely believes that she wants to join the reading group on *Being and Time* because she has a

distorted view of the sort of person she is. She isn't really an excessively sophisticated intellectual with a taste for Teutonic obscurity, but a more straightforward plain-spoken type who prefers *A Materialist Theory of the Mind*.

That we have privileged access to our thoughts has some experimental support. Subjects have been instructed to "think aloud" while performing a variety of cognitive tasks "in their heads": multiplying numbers, solving the tower of Hanoi puzzle, counting the windows in their living rooms, remembering and recalling a matrix of digits, and so on (Ericsson and Simon 1993). The subjects' self-reports can then be checked against theoretical models of problem solving, the subjects' response latencies, and so on. Summarizing, Nichols and Stich write that:

> both commonsense and experimental studies confirm that people can sit quietly, exhibiting next to no overt behaviour, and give detailed, accurate self-reports about their mental states. (Nichols and Stich 2003: 158)

The claim of *privileged* access to our own thoughts is a watered-down version of what is sometimes called *infallible*, *incorrigible*, or *indubitable* access:

IN Necessarily, if S believes she is in M, then she is in M/knows that she is in M

What is sometimes called *self-intimating* access is a near-converse of IN:

S-I Necessarily, if S is in M, she knows/is in a position to know that she is in M[2]

Privileged access should not be confused with S-I and its watered-down variants, which have little to recommend them, at least unless highly qualified and restricted. Post-Freud, the idea that there are subterranean mental currents not readily accessible to the subject is unexceptionable, even if Freud's own account of them is not.

Privileged access is the first part of McKinsey's "philosophical commonplace". The second part is that we have *peculiar* access to our mental states: as McKinsey says, we know about them "in a . . . way that is available to no one else". Kylie knows that she is thinking that trees die without water by "introspection", but whatever introspection is, she cannot employ it to discover that someone *else* is thinking that trees die without water. In McKinsey's preferred terminology, Kylie can know "a priori" that she is thinking that trees die without water. (This terminology is hardly apt, since one leading theory of self-knowledge classifies it as a variety of *perceptual* knowledge: see §III.)

It is important to distinguish privileged and peculiar access because they can come apart in both directions. Consider Ryle, who holds that we have access to our own minds in the same way that we have access to others' minds—by observing behavior—and thus denies that we have *peculiar* access. "The sorts of things that I can find out about myself are the same as the sorts of things that I can find out about other people,

[2] For the belief-versions of both IN and S-I see Armstrong (1968: 101); as Armstrong notes, the term 'self-intimation' is due to Ryle. In the terminology of Williamson (2000: 95), the knowledge-version of S-I is the thesis that the condition that one is in M is "luminous".

and *the methods of finding them out are much the same*" (Ryle 1949: 149, emphasis added). Yet he thinks that we have *privileged* access to our mental states (at least sometimes), because we (sometimes) have better behavioral evidence about ourselves—greater "supplies of the requisite data" (149).[3]

Ryle's position shows why privileged access need not be peculiar. For an illustration of the converse, suppose that one often mistakes beeches for elms by sight. Often one's belief that one sees an elm is mistaken, or at any rate not knowledge. One's access to one's mental state of *seeing an elm* is nothing to write home about: much knowledge of one's enviroment and of others' mental states is considerably easier to atttain. Yet one has peculiar access to the fact that one sees an elm: one does not need the usual third person evidence—that one's eyes are open and pointing towards an elm, one has an unobstructed view of the elm, the lighting is good, and so on.

It should be emphasized that the alleged conflict between self-knowledge and externalism has its source in *peculiar* access, not *privileged* access. Unless privileged access is taken to be something like *infallible* access, there is not even the appearance of a conflict with externalism. Peculiar access to the existence of water (for example) is the issue. McKinsey's argument gives rise, as Davies says, to the puzzle of *armchair* knowledge.

III. Inner-sense

According to Armstrong:

> Kant suggested the correct way of thinking about introspection when he spoke of the awareness of our own mental states as the operation of 'inner sense'. He took sense perception as the model of introspection. By sense-perception we become aware of current physical happenings in our environment and our body. By inner sense we become aware of current happenings in our own mind. (Armstrong 1968: 95; see also Armstrong 1981, Lycan 1987: ch. 6, 1996: ch. 2, Nichols and Stich 2003: 160–4)

By using her outer eye, Kylie knows that there is a lot of water in Lake Burley Griffin; according to the inner-sense theory, by using her inner eye, Kylie knows that she is thinking of water.

The inner-sense theory does offer a nice explanation of *peculiar* access: for obvious architectural reasons, the (presumably neural) mechanism of inner-sense is only sensitive to the subject's own mental states. In exactly the same style, our faculty of

[3] The claim of "Privileged Access", in Ryle's sense, is this: "(1) . . . a mind cannot help being constantly aware of all the supposed occupants of its private stage, and (2) . . . it can also deliberately scrutinize by a species of non-sensuous perception at least some of its own states and operations" (Ryle 1949: 158). In the contemporary literature, 'privileged access' is often used approximately for what (in the text) is described as privileged *and* peculiar access (see e.g. Alston 1971; Moran 2001: 9–10). Ryle's characterization of "non-sensuous perception" denies that there are "any counterparts to deafness, astigmatism, . . ." (1949: 157); this means that Ryle's Privileged Access implies but is not implied by the conjunction of privileged and peculiar access (in the sense of the text).

proprioception explains the "peculiar access" we have to the position of our own limbs (see Armstrong 1968: 307).

But it does not explain *privileged* access. In fact, it leaves it something of a mystery. Why is inner-sense less prone to error than the outer senses? Why are (as we commonly suppose) our opinions about our own desires more error-prone than those about our thoughts? Are desires located in some dark and obscure place in the brain?

And finally, if there is an inner-sense, then presumably it might be damaged or absent, while sparing the rest of a normal subject's cognitive capacities. In other words, the condition Shoemaker (1988) calls "self-blindness" could occur, if the inner-sense theory is correct. Perhaps self-blindness *is* (metaphysically) possible, but no *actual* psychological condition seems to come close.[4]

IV. Economy and extravagance

On one (minority) view of our knowledge of language, it requires nothing more than a general purpose learning mechanism. On the alternative Chomskean picture, it requires a dedicated faculty, a "language organ". The minority view is an *economical* account of our linguistic knowledge: no special purpose epistemic capacities are required. The Chomskean view, on the other hand, is *extravagant*: given the meager input—the "poverty of the stimulus"—the general purpose mechanism is supposed incapable of generating the required torrential output.

A similar distinction can be drawn for self-knowledge. Let us say that a theory of self-knowledge is *economical* just in case it explains self-knowledge solely in terms of epistemic capacities and abilities that are needed for knowledge of other subject matters; otherwise it is *extravagant*.[5] Ryleanism is economical: the capacities for *self*-knowledge are precisely the capacities for knowledge of the minds of *others*.[6] The theory defended in Shoemaker (1994) is also economical: here the relevant capacities are "normal intelligence, rationality, and conceptual capacity" (236). The inner-sense

[4] According to Nichols and Stich, "certain kinds of schizophrenia might involve damage to the Monitoring Mechanism [their term for an Armstrongian faculty of inner-sense] that does not affect other components of the mind-reading system" (2003: 190). Their main examples concern "passivity symptoms", such as delusions of control (in which patients attribute their own actions to other agents). Nichols and Stich also cite Hurlburt's work in "experience sampling" in support of the view that some schizophrenic patients have difficulty reporting their imagery and thoughts when symptomatic (190–1; cf. Hurlburt 1990: esp. 204–8, 224–5, 238–9). However, none of these cases remotely approximate Shoemakerian self-blindness. Schizophrenic self-blindness is limited, and anyway schizophrenia patients suffer from other cognitive impairments.

[5] A qualification: knowledge of the "other subject matters" should not itself require mental evidence. On one traditional view (Locke's, for example), *all* empirical knowledge is founded on mental evidence about perceptual appearances; on that view (assumed false here), the correct theory of self-knowledge is extravagant, not economical.

[6] Carruthers (2009) is a recent defense of a sophisticated view of this type. Like the account to be defended later, inner speech plays an important role, but Carruthers denies that his account is an extension of the Evans-style transparency procedure mentioned a few paragraphs below.

theory, on the other hand, is *extravagant*: the organs of outer perception, our general capacity for rationality, and so forth, do not account for all our self-knowledge—for that, an additional mechanism, an "inner eye", is needed.

The inner-sense theory has no explanation of privileged access. Ryleanism implausibly denies that we have peculiar access. On Shoemaker's theory, "there is a conceptual, constitutive connection between the existence of certain sorts of mental entities and their introspective accessibility" (Shoemaker 1994: 225), but Shoemaker's case for this very strong claim is open to question.[7]

However, in the case of belief there *is* an economical theory that is (a) Shoemakerian in spirit, and that (b) (arguably) explains *both* privileged and peculiar access.

Consider the following well-known passage from Evans:

> [I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me "Do you think there is going to be a third world war?," I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?". (Evans 1982: 225)[8]

One apparently finds out that one believes that it's raining by determining whether it's raining: knowledge that one has this belief, insofar as it rests on perceptual evidence at all, rests on perceptual evidence about the weather, not on perceptual evidence of one's behavior or anything mental. That is, one *reasons* from the evidence that it's raining, to the conclusion that one believes that it's raining.

Of course, this is not to say that such reasoning is *good* reasoning. In fact, prima facie it is terrible reasoning. Ignoring this difficulty, the important point is that the particular capacity to reason from the evidence that p to the conclusion that one believes that p does not involve any special mechanism of inner-sense—given a stock of mental concepts, it rides on the coattails of the general capacity for reasoning.

The next section briefly argues for two claims. First, that this Evans-style "transparency" procedure is indeed good reasoning, and so can yield self-knowledge. Second, that it can also explain privileged and peculiar access. The rest of the paper tries to extend this account to the case of thought.

V. An epistemic rule for belief[9]

Consider a very simple example of reasoning: Kylie hears the kettle whistle, and concludes that the water is boiling. By hearing that the kettle is whistling, Kylie knows that the kettle is whistling; by reasoning, she knows that the water is boiling.

[7] See Byrne (2005: 82–92).

[8] See also Dretske (1994, 1995) and Gordon (1996). A similar view can be found in Husserl; see Thomasson (2003) for an interesting discussion.

[9] See also Byrne (2005: 93–5).

We can usefully think of Kylie acquiring knowledge of the water's boiling by following a recipe or rule. If we say that an *epistemic rule* is a conditional of the following form:

R If conditions C obtain, believe that p[10]

then the epistemic rule that Kylie follows is:

KETTLE If the kettle whistles, believe that the water is boiling

What does it mean to say that Kylie *follows* this rule on a particular occasion? The semi-stipulative answer is this: Kylie believes that the water is boiling *because* she recognizes that the kettle is whistling. The 'because' is intended to mark the kind of reason-giving causal connection that is often discussed under the rubric of 'the basing relation'. Kylie might recognize that the kettle is whistling, and believe that the water is boiling for some *other* reason: in this case, she does not form her belief because she recognizes that the kettle is whistling.

So S follows the rule R ('If conditions C obtain, believe that p') on a particular occasion iff on that occasion:

(i) S believes that p because she recognizes that conditions C obtain

which implies:

(ii) S recognizes (hence knows) that conditions C obtain
(iii) conditions C obtain
(iv) S believes that p

Following KETTLE tends to produce knowledge about the condition of the water (or so we may suppose), and hence it is a *good* rule. Following *bad* rules tends to produce false and unjustified beliefs, for example:

ASTROLOGY If the daily horoscope predicts that p, believe that p

ASTROLOGY is also an example of a *schematic* rule. One *follows* a schematic rule just in case one follows a rule that is an instance of the schematic rule; a schematic rule is *good* to the extent that its instances are.

If the antecedent conditions C of an epistemic rule R are not specified in terms of the rule follower's mental states, R is *neutral*. A schematic rule is neutral just in case some of its instances are. Thus, the claim that S can follow a neutral rule does not presuppose that S has the capacity for self-knowledge. KETTLE and ASTROLOGY are neutral rules; 'If the type looks blurry, believe that you need glasses' is not.[11]

[10] It should be emphasized that the linguistic formulation of the rule *only* plays a heuristic role—*all* the work is done by the account of *following* a rule (see immediately below).

[11] 'You' refers to the rule-follower; tenses are to be interpreted so that the time the rule is followed counts as the present.

Self-knowledge is our topic, not skepticism: knowledge of one's environment (including others' actions and mental states) and reasoning (specifically, rule-following of the kind just sketched) can be taken for granted. So, in the present context, it is not in dispute that we follow neutral rules, including neutral rules with mentalistic fillings for '*p*', like 'If S blushes, believe that S is embarrassed'; neither is it in dispute that some neutral rules are good rules.

Evans's observation in the passage quoted in the previous section may be recast using the apparatus of epistemic rules as follows. Knowledge of one's beliefs may be obtained by following the neutral schematic rule:

BEL If *p*, believe that you believe that *p*

Since the antecedent of BEL expresses the content of the mental state that the rule-follower ends up believing she is in, BEL can be called a *transparent* rule.

One is only in a position to follow BEL by believing that one believes that *p* when one has recognized that *p*. And recognizing that *p* is (inter alia) coming to *believe* that *p*. BEL is *self-verifying* in this sense: if it is followed, the resulting second-order belief is true. Compare a third-person version of BEL:

BEL-3 If *p*, believe that Kylie believes that *p*

BEL-3 is of course not self-verifying: the result of following it may be (indeed, is very likely to be) a false belief about Kylie's beliefs.

Say that S *tries* to follow rule R iff S believes that *p* because S *believes* that conditions C obtain. That S follows R entails that she tries to follow R, but not conversely. If one tries to follow KETTLE but does not succeed, then one will not *know* that the water is boiling; if one's belief about the water is true, that is just an accident.

Sometimes one will not succeed in following BEL; instead one will merely try to follow it. That is: one believes but does not know that *p*, and thereby concludes that one believes that *p*. Here one's second order belief that one believes that *p* will be *true*. Since *trying* to follow BEL cannot lead to error, BEL can be called *strongly* self-verifying.[12] Hence (we may fairly suppose) this situation will be commonplace: trying to follow BEL, one investigates whether *p*, *mistakenly* concludes that *p*, and thereby comes to *know* that one believes that *p*. In these cases, one will know that one believes that *p* on the basis of no evidence at all.[13]

[12] An example of a rule that is self-verifying but not strongly so is:

KNOW If *p*, believe that you know that *p*

[13] One might have the following worry at this point. Haven't we learnt from Gettier cases (Gettier 1963) that knowledge cannot be based on reasoning through a false step? So if Kylie infers that she believes that it's raining from the false premise that it's raining, how can she know that she believes it's raining? The short answer is that a better diagnosis of the Gettier examples is that *safety* (in the sense of Sosa 1999 and Williamson 2000: ch. 5) is a necessary condition for knowledge, not that no reasoning through false steps is a necessary condition for knowledge. And beliefs produced by trying to follow BEL will often be safe (see Byrne 2005:

Given that we follow rules like KETTLE, it should not be in dispute that we *can* follow BEL. Given the plausibility of Evans's observation about the procedure we actually follow, it should not be in dispute that we *do* follow BEL. BEL offers an obvious explanation of *peculiar* access: as just noted, BEL-3 is a very bad rule indeed. Privileged access is explained—at least on the face of it—because BEL is strongly self-verifying.

The account just sketched is not a version of the inner-sense theory. It is *economical*, like behaviorism. Taking the capacity to follow good neutral rules for granted, knowledge of what one believes comes along more-or-less for free. Since this capacity belongs to the department of reasoning, not perceiving, Shoemaker's idea that the source of self-knowledge can be traced to "rationality" is vindicated, albeit not via his preferred route.

VI. Silent soliloquy

Much of our thinking "occurs in inner speech", or in what Ryle calls an "internal monologue or silent soliloquy" (Ryle 1949: 28). In some sense—yet to be explained—one hears oneself thinking, with "the mind's ear". A natural idea, then, is that Kylie knows that she is thinking about water, and that she is thinking that trees die without water, because she eavesdrops on herself uttering, in the silent Cartesian theater, 'Trees die without water'. At any rate, that might be one way Kylie can know that this is what she is thinking. The rest of this paper pursues this suggestion. The project of leveraging this into an *economical* account of self-knowledge might initially seem hopeless—reassurance will be offered several sections later.

The paradoxical-seeming claims of the previous paragraph need to be cleared up first. When one engages in silent soliloquy, there are no sounds in one's head—that's why the soliloquy is silent. But, if there really is inner *speech*, there are sequences of phonemes, and so sounds. Hence there is no such thing as inner speech.

Instead, speech is (phonologically) *represented*. When Kylie "hears" her "internal monologue", she is in a quasi-perceptual state that represents an utterance of the sentence 'Trees die without water'. (This claim will receive some defense later, in §IX.)

There is nothing of which Kylie is aware when she "hears" her inner speech. A fortiori, she is not aware of any episode of thinking about water. Still, it seems plausible that this appearance of an inner monologue enables Kylie to know that she is thinking about water. Why should this be so?

(*Thinking about* x is the subject of the next five sections; *thinking that* p will be treated in §XII.)

96–8). A similar point is crucial for the following account of the epistemology of thought, which always involves reasoning through a false step. (Thanks here to Benj Hellie.)

VII. Outer and inner speech

Ryle noted that an important source of information about others is provided by their "unstudied talk", utterances that are "spontaneous, frank, and unprepared" (Ryle 1949: 173). Chatting with Kylie over a few beers is the best way of discovering what she believes, wants, and intends. "Studied" talk, on the other hand, is not so revealing. If Kylie is a politician defending the federal government's policies on water, she might assert that the water shortage will soon be over without believing it will be.

However, in the umbrella sense of "thinking about *x*" with which we are concerned, both unstudied and studied talk provide excellent evidence about the utterer's thoughts. Even if the Hon. Kylie, MP, doesn't *believe* that the water shortage will soon be over, she was presumably thinking *about* water. Outer speech on such-and-such topic is produced by mental activity about that same topic.

If someone outwardly utters 'The water shortage will soon be over' then (usually) she says something, namely that the water shortage will soon be over. She says—and so thinks—something *about* water. Does the same point hold for inner speech? An affirmative answer is not trivial, because the production of inner speech, unlike the production of outer speech, might have nothing to do with the *semantics* of the words. Perhaps an inward utterance of 'The water shortage will soon be over' is produced in a similar semantics-insensitive manner as the inward utterance of 'Dum diddley', or other meaningless string.

But this possibility can be dismissed by noting that outer speech and inner speech often perform the same function, moreover one for which the semantics of the words is crucial. One may cajole or encourage oneself out loud; one may also do so silently. Inner speech and outer speech may be seamlessly interleaved in a conversation. One may recite a shopping list out loud to preserve it in working memory; silent recitation will do just as well.[14]

By "hearing" herself inwardly utter 'Trees die without water', then, Kylie can know that she is thinking about water. As Ryle puts it:

> We eavesdrop on our own voiced utterances and our own silent monologues. In noticing these we are preparing ourselves to do something new, namely to describe the frames of mind which these utterances disclose. (Ryle 1949: 176)

When Kylie utters (out loud) 'Trees die without water', there is something she is aware of—the utterance of that sentence. She is not thereby aware *of* a mental episode of thinking about water, although she can thereby be aware *that* she is thinking about water. When Kyle utters in "silent soliloquy", 'Trees die without water', there is *nothing* she is aware of, a fortiori she is not aware of a mental episode of thinking about water.

[14] On one standard model of working memory, this involves the so-called "inner ear", a short-term memory store for phonological information which is part of the "phonological/articulatory loop" (Baddeley 1986: ch. 5).

And even if there *were* an "inner utterance", occuring in some ethereal medium, it would still be wrong to say that Kylie is aware of any mental episode. The outer utterance is not itself an episode of thinking, but something produced by such an episode; likewise, if there were (per impossibile) an inner utterance it wouldn't be an episode of thinking either.

According to Carruthers:

> Introspection informs us, in fact, that many of our thoughts are expressed entirely in natural language . . . I can report that most of my thoughts occur in the form of imaged conversations . . . (Carruthers 1996: 50–1)

But surely introspection informs us of no such thing. There is no inner voice, and there are no inner utterances of English words. Introspection does not inform us that thoughts are represented in natural language, any more than it informs us that visual imagination involves pictorial inner representations. Introspection does not count against Fodor's claim that we think in Mentalese.

Let us recast the main point of this section using the apparatus of epistemic rules. One can know what one thinks by hearing one's outer speech. Likewise, one can know what one thinks by "hearing" one's inner speech. One route to what one is thinking is the rule:

THINK-ALOUD If you outwardly speak about x, believe that you are thinking about x

But a much more common route, and the route of present interest, does not require using one's ears:

THINK* If you inwardly speak about x, believe that you are thinking about x

The argument so far is simply that THINK* is a good rule. Nothing has been said yet about privileged or peculiar access. Moreover, THINK* is not a *neutral* rule. Since "inwardly speaking about x" is auditorily imagining—perhaps better, *phonologically* imagining—words that are about x, THINK* could be rephrased as follows:

THINK** If you auditorily imagine words that are about x, believe that you are thinking about x[15]

THINK**, and hence THINK*, are obviously not neutral—their antecedents concern the rule-follower's mental states. If THINK* is the best that can be done, then there is no economical theory of the epistemology of thought.

[15] For reasons why 'phonologically imagine' might be a more accurate expression than 'auditorily imagine', see Mackay (1992: 127–30). But the latter expression is more familiar, and will be used throughout. It should be emphasized that different sorts of 'auditory imagination' (inner speech, imagining the speech of another person, imagining noises, imagining music) may well involve distinct cognitive systems.

VIII. Downgrading to economy

When one speaks aloud, one is in a position to know that one is speaking. The fact one is in a position to know is a fact about oneself. One is also in a position to know the impersonal fact, that some speaking (perhaps not even by a speaker) is occurring. To know the latter, one just needs to identify speech; one does not need any capacity to know one's communicative intentions, or laryngeal or jaw movements. That is, at least in principle one may know that some speaking—which is, in fact, one's own speech—is occurring, without being in a position to know that this is one's own speech.

A similar point holds for inner speech—with one important caveat to be mentioned in the next paragraph. When one auditorily imagines words, one is in a position to know that one is auditorily imagining words. The fact one is in a position to know is a fact about oneself, specifically a fact about one's mental activity. Pretend—just for the remainder of this paragraph—that the imagined words are actually produced in a cranial concert hall, and heard by an inner ear; scientists can detect inner speech by attaching sensitive microphones to the head. Then one is also in a position to know the impersonal fact, that some inner speech (perhaps not even by a speaker) is occurring. (Speech in the cranial concert hall sounds quite distinctive.) An Evans-style "transparency" observation applies here: to know that inner speech is occurring, one just needs to identify inner speech. And this does not require the capacity for self-knowledge, for instance the capacity to know that one is *imagining* speech.

Here is the caveat. One cannot *know* that inner speech is occurring, because there isn't any. Similarly, if one forms a visual image of a purple kangaroo, one cannot know that there is a visual image of a purple kangaroo—something closely resembling a real picture of a purple kangaroo—because there is no such image.

(Of course, visually imagining a purple kangaroo might involve the manipulation of a picture-like representation in the brain. But this picture-like representation is (a) not something that one is aware of and (b) not much like a picture of a purple kangaroo—it won't be purple, for one thing.[16])

Although one cannot *know* that there is an image of a purple kangaroo, one can be in a related state. One can simply take the appearances at face value, and *believe* that there is a visual image of a purple kangaroo. And presumably this is the inclination of many of us, at least.

(This inclination may be part of the explanation of the sense datum theory's appeal: if to imagine a purple kangaroo is to be aware of an image of a purple kangaroo, then, since seeing introspectively resembles visually imagining, seeing must also involve being aware of an image—presumably an especially vivid one.)

The capacity to believe that there is an image of a purple kangaroo whenever one visually imagines a kangaroo is not a capacity that presupposes the capacity for

[16] Notoriously, pictorialists about mental imagery are often tempted to deny (a): a recent example is Kosslyn et al. (2006: 48–9).

self-knowledge; in particular, one may have the former capacity without having the capacity to know that one is imagining a purple kangaroo.

Likewise, the capacity to believe that inner speech is occurring whenever one auditorily imagines words does not presuppose the capacity for self-knowledge. So, although THINK★ is not neutral, a related neutral rule is, namely:

THINK If the inner voice speaks about *x*, believe that you are thinking about *x*

If one follows THINK, one recognizes, hence knows, that the inner voice speaks about *x*. Since there is no inner voice, there is no such knowledge to be had, and one cannot *follow* THINK. However, one can *try* to follow THINK (see §V above).

IX. Telling the inner voice from the outer voice

One *can* tell the inner voice apart from the outer voice. But nothing has yet been said about why this is so.

§VI claimed without argument that inner speech involves representing an utterance, akin to the way in which utterances are represented when outer speech is heard. A lot of the experimental evidence concerns the corresponding thesis for visual imagination: visually imagining involves representing an arrangement of objects, akin to the way in which an arrangement of objects is represented when one sees the scene before one's eyes. Since the visual thesis lends plausibility to the inner speech thesis, the latter can be indirectly supported by citing the evidence for the former.[17]

First, seeing and visually imagining overlap substantially at the neural level. A recent functional magnetic resonance imaging (fMRI) study concluded that "visual imagery and visual perception draw on most of the same neural machinery" (Ganis et al. 2004: 226).[18]

Second, there are many interference effects between visual imagery and visual perception, which are not as pronounced between (say) visual imagery and auditory perception in different modalities (Kosslyn 1994: 54–8).

Third, there is the phenomenon of eidetic imagery, in which subjects report their images to be very much like photographs (Haber 1979).

Fourth, perception can be mistaken for visual imagery (the Perky effect). Asked to imagine a red tomato while staring at a white screen, subjects will often fail to notice a clearly visible red round image subsequently projected onto the screen, and will take its size and shape to be features of the imagined tomato.[19]

Fifth, visual imagery can (arguably) be mistaken for perception (the reverse-Perky effect). For instance, a cortically blind subject (H.S.) with spared visual imagery denied

[17] For a summary of the direct evidence see Mackay (1992).

[18] For a similar study for musical imagery, see Zatorre and Halpern (2005).

[19] Hume basically noted the Perky effect at the beginning of the *Treatise*: "it sometimes happens, that our impressions are so faint and low, that we cannot distinguish them from our ideas" (Hume 1740/1978: 2).

she was blind (Anton's syndrome), which may have "resulted from a confusion of mental visual images with real percepts" (Goldenberg et al. 1995: 1373):

H.S. was given a comb and recognized it from touch . . .
G.G.: Are you really seeing it, or is it only a mental image?
H.S.: I think I am seeing it a little, very weakly . . .
G.G.: What does "weakly" mean?
H.S.: It is vague and . . . somehow farther away, blurred. (Goldenberg et al. 1995: 1378)[20]

Sixth, and finally, we ordinarily and naturally speak of visual *images* and of an inner *voice*. (And not so frequently, incidentally, of "inner smells".) This is straightforwardly explained by the hypothesis that visual and auditory imagery share a distinctive sort of representational content with, respectively, vision and audition.

It might be objected that if auditory imagery is a kind of quasi-hearing, then this will turn auditory imagery into auditory hallucination. And if the inner voice is *hallucinated*, then how can we so effortlessly tell that the inner voice is not outer?

But this can be explained by assuming that the representational content of imagery is seriously degraded, compared to the representational content of perception. (Sometimes it is not—that is when imagery crosses over into hallucination.) As is often pointed out, imagery is indeterminate in a way that perception typically isn't. For example, one can visually imagine a striped tiger without, for some n, imagining a tiger with n stripes, and one can auditorily imagine speech without imagining that the voice is loud or soft, masculine or feminine, near or far.

Hume notoriously claimed that "ideas" (images) differ only from "impressions" (sense data) in "their degree of force and vivacity" (Hume 1740/1978: 2). Ideas, on Hume's view, are like faded photographs before the mind's eye, or sounds from worn cassette tapes played to the mind's ear. But even after ideas and impressions are rejected as fictions, an insight can be salvaged from Hume's suggestion: imagery and perception differ in their degree of information.

X. Privileged and peculiar access again

The supposition that we try to follow THINK can explain privileged and peculiar access. Peculiar access is explained because "I cannot overhear your silent colloquies with yourself" (Ryle 1949: 176). Typically one's "inner voice" will be entirely self-generated. Admittedly, sometimes one's inner voice might speak about *x* because one overhears someone else speaking (and so thinking) about *x*. However, if in such circumstances one concluded from the pronouncements of the inner voice that Kylie (say) is thinking about *x*, this would at best be an accidentally true belief, not knowledge.

[20] For the reverse-Perky effect at the level of memory, see Intraub and Hoffman (1992).

What about privileged access? Consider the rule:

THINK^{K} If Kylie speaks about *x*, believe that Kylie is thinking about *x*

This is a good rule, as Ryle in effect observed. Suppose one follows THINK^{K} on a particular occasion: Kylie utters 'Trees die without water', one recognizes (hence knows) that Kylie is speaking about water, and thereby concludes that Kylie is thinking about water. Let us grant that if Kylie is speaking about water then she is thinking about water. One's belief that Kylie is thinking about water will then be true and (at least in a typical case) will also be knowledge.

If one adopts the policy of following THINK^{K}, one will not always succeed. In particular, sometimes one will merely try to follow it. For instance, perhaps one misidentifies the speaker as Kylie, or mishears her; the resulting belief about Kylie's thoughts will not then be knowledge. The policy of following THINK^{K} might well not produce, in perfectly ordinary situations, knowledge of what Kylie thinks.

For comparison, suppose that Kylie adopts the policy of following THINK. If (as Kylie would put it) the inner voice utters 'Trees die without water', Kylie thereby concludes that she is thinking about water. Kylie's policy can never succeed: her belief about the inner voice's pronouncements are always false.

Although Kylie can never follow THINK, but only *try* to follow it, for the purposes of attaining self-knowledge that doesn't matter. If she believes that she hears the inner voice utter 'Trees die without water', it is not flatly impossible that she has not auditorily imagined that sentence, but it is very unlikely. And even if she has mistaken an outer voice for her inner one (an example of the auditory Perky effect), her belief that she is thinking about water is *still* true, and will likely be knowledge. (If you hear someone else speak about water, the very act of understanding their words requires that you think about water.)

The beliefs produced by trying to follow THINK are not absolutely guaranteed to be true (here THINK differs from the rule BEL introduced in §V). A fortiori, they are not absolutely guaranteed to be knowledge. But they are very likely to be; much more so than the beliefs produced by following third-person rules like THINK^{K}. That kind of epistemic access may be privileged enough.[21]

[21] A notoriously puzzling symptom of schizophrenia is "thought insertion" (one of the "passivity symptoms" mentioned in note 4). Patients claim that certain thoughts are not their own, despite being "in their minds". On the present view of the epistemology of thought, such patients "hear" the inner voice speak about *x*, but do not (try to) follow THINK, and conclude that they are thinking about *x*. Instead, they conclude (paradoxically) that although there is a thought about *x* in their minds, they are not the thinker of that thought.

Thought insertion is too complicated to discuss here, but it should be noted that the present view does fit nicely with accounts of thought insertion on which patients attribute their own inner speech to an external agent. See e.g. Jones and Fernyhough (2007), and (for an even better fit) Langland-Hassan (2008).

XI. A complication: THINK for philosophers

It is probably safe to say that many ordinary people believe that there is an inner voice (in some unspecified sense importantly like an outer voice) that they can hear with an inner ear (in some unspecified sense importantly like an outer ear). Since that is just how things seem, what else are they supposed to think?

But what of the enlightened few who are of the contrary opinion? They apparently can't try to follow THINK, so does that mean that their enlightenment comes at the steep price of self-ignorance? Or will the enlightened—unlike the vulgar—need a faculty of inner sense to find out what they are thinking?

However, it is not an immediate consequence that the enlightened cannot try to follow THINK. What is an immediate consequence is that if they do try to follow THINK, and hence believe that there is an inner voice, then their beliefs are inconsistent. The belief that there is an inner voice plays a highly circumscribed role, confined to inferences of the THINK variety; their belief that there is no inner voice plays a much more expansive role, of the sort associated with paradigmatic cases of belief. When trying to follow THINK, the enlightened would briefly allow their belief that there is an inner voice, normally firmly suppressed, to control their reasoning.

On this view, imagery is not contingently connected to belief; rather, it is constitutively involved with belief. Is there any reason to suppose this is so, other than to save the present theory? If *perception* constitutively involves belief, then given the close kinship between perception and imagery, the parallel claim for imagery should not seem implausible. Moreover, the idea that perception is, at the very least, constitutively connected to a *tendency* to believe is very appealing. As A. D. Smith puts it, "How can I disbelieve my senses *if I have nothing else to go on?*" (2001: 289).

The phenomenon that is widely taken to block the stronger conclusion that perception constitutively involves, not a mere tendency to believe, but belief itself, is the phenomenon of known illusion. It visually appears that the lines in the Müller-Lyer figure are unequal, but one plainly believes that they are equal. Hence—it is commonly argued—the "content of perception", the proposition that the lines are unequal, need not be believed.

Consider the alternative hypothesis that the belief that the lines are unequal is present, but inhibited by the contrary belief that the lines are equal. This need not involve the postulation of novel psychological mechanisms. First, inconsistent beliefs are commonplace. Second, certain delusional beliefs (as in the Capgras delusion, for instance) can be largely disconnected from the rest of the subject's world view: despite believing that their spouse (say) has been replaced by an imposter, "[m]ost Capgras patients do not show much concern about what has happened to the real relatives; they do not search for them, or report their disappearance to the police" (Stone and Young 1997: 333). On the alternative hypothesis, when one knowingly looks at the Müller-Lyer figure one's belief that the lines are unequal is like an inferentially isolated delusion. One consideration in favor of this hypothesis is that it nicely explains the

fact that, absent any reason for thinking otherwise, one will believe—in the usual inferentially promiscuous way—that the lines are unequal (Byrne 2009: 450–1). In contrast, the orthodox alternative has no explanation at all.[22]

So the view that imagery constitutively involves belief—harmless delusions about a shadowy inner world of (at least) visual and auditory images—is not as unmotivated as it may seem. A preliminary case has been made that THINK requires no modification to accommodate the self-knowledge of the enlightened.

XII. Extensions: Imagistic thinking, thinking that p

Much of our thinking "occurs in pictures" (as we say), not in words. The extension of THINK to this case presents no further problem:

THINK-IMAGE If the inner picture is about *x*, believe that you are thinking about *x*

Knowledge that one is thinking that *p*—that trees die without water, say—can be accommodated by combining THINK with the rule BEL from §5:

THINK-THAT If the inner voice says that *p* and *p*, believe that you are thinking that *p*

It is worth noting that the present account resolves a puzzle about thought. We often say we are thinking *of x* without being able to come up with *any* property that the thought predicates of *x*. That is arguably a unique feature that distinguishes thought from belief, desire, and intention.

Consider an example. One is thinking of Barry Humphries, say. Suppose one knows this because one ostensibly discerns a visual image of Humphries. Although one might not be thinking *that* Humphries is F, there *is* a reportable predicational component to the thought—Humphries is imagined dressed as Dame Edna, say, or as having black hair. But sometimes one can think of Humphries with no apparent predicational component at all—one is *simply* "thinking of Humphries" (cf. "Think of a number"). The explanation is that one may *simply* auditorily imagine the name 'Barry Humphries' (or—in the case of "thinking of a number"—the numeral '7').

XIII. Unsymbolized thinking?

Suppose the account just outlined is right. We can, and do, know what we think by trying to follow THINK. Privileged and peculiar access are explained, and our general capacity for reasoning suffices for self-knowledge. With that all granted, there might still be rain on the parade.

[22] Further support for the constitutive hypothesis can be found in Smith (2001) and Craig (1976). For a defense of the view that delusory "beliefs" are the genuine article, see Stone and Young (1997: 351–9), and Bayne and Pacherie (2005).

In "thought sampling", subjects are equipped with a beeper that sounds at random intervals. They are instructed to "'freeze' their ongoing experience and write a description of it in a notebook" (Hurlburt 1993: 10). Some subjects report what Hurlburt calls "Unsymbolized Thinking"—"imageless thought", in more familiar terminology from the eponymous controversy in early twentieth-century psychology. Unsymbolized thinking is:

> the experience of an inner process which is clearly a thought and which has a clear meaning, but which seems to take place without symbols of any kind, that is, without words, images, bodily sensations, etc. (1993: 5; see also Hurlburt and Akhter 2008)

If subjects can "just know" (5) that they are thinking about *x* without noting the (apparent) presence of words or images about *x*, then this fact has not yet been explained.

One explanation might be that subjects are doing some quick self-interpretation: given my behavior, circumstances, and/or other topics of thought, it is to be expected that I am thinking about *x*. (This hypothesis is argued for in some detail in Carruthers 1996: 239–44.) Alternatively, perhaps inner speech and visual imagery do provide the necessary basis for inference along the lines of THINK, but subjects are reporting the content without reporting (perhaps because they have forgotten) the speech or imagery itself. As when one reports that such-and-such without reporting whether one heard someone assert it or whether one read it in the newspaper, the subjects are reporting the message but not the medium.

Hurlburt has recently clarified his view of unsymbolized thinking, with the result that an explanation along the latter lines seems the more plausible, at least for some cases. As far as his data show, imagery and inner speech are not entirely absent in unsymbolized thinking, but are rather diminished in relevance and are not at the focus of attention:

> Unsymbolized thinking . . . may well have some kind of (probably subtle) sensory presentation . . .
> . . . The apprehension of an unsymbolized thought may involve the apprehension of some sensory bits, so long as those sensory bits are not organized into a coherent, central, thematized sensory awareness. (Hurlburt 2009: 149–50)

Provided that the "sensory bits" provide enough clues, which Hurlburt does not deny, then unsymbolized thinking does not present a problem.

Of course, if there is a way of apprehending one's thoughts "directly", which explains both peculiar and privileged access, then there's no need to try to explain knowledge of unsymbolized thinking using the present model. However, despite the great weight placed on the epistemology of thought in the Cartesian tradition, plausible accounts of how we know that we are thinking are remarkably thin on the ground.

References

Alston, W. P. (1971), "Varieties of Privileged Access", *American Philosophical Quarterly* 8: 223–41. Page reference to the reprinting in Alston (1989).

——(1989), *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca, NY: Cornell University Press).

Armstrong, D. M. (1968), *A Materialist Theory of the Mind* (London: Routledge and Kegan Paul).

——(1981), "What Is Consciousness?", *The Nature of Mind and Other Essays* (Ithaca: Cornell University Press).

Baddeley, A. (1986), *Working Memory* (Oxford: Oxford University Press).

Bayne, T., and Pacherie, E. (2005), "In Defence of the Doxastic Conception of Delusions", *Mind and Language* 20: 163–88.

Byrne A. (2005), "Introspection", *Philosophical Topics* 33: 79–104.

——(2009), "Experience and Content", *Philosophical Quarterly* 59: 429–51.

Carruthers, P. (1996), *Language, Thought and Consciousness: An Essay in Philosophical Psychology* (Cambridge: Cambridge University Press).

——(2009), "How We Know Our Own Minds: The Relationship between Mindreading and Metacognition", *Behavioral and Brain Sciences* 32: 121–82.

Craig, E. (1976), "Sensory Experience and the Foundations of Knowledge", *Synthese* 33: 1–24.

Davies, M. (2000), "Externalism and Armchair Knowledge", in P. Boghossian and C. Peacocke (eds), *New Essays on the A Priori*. (Oxford: Oxford University Press).

Dretske, F. (1994), "Introspection", *Proceeding of the Aristotelian Society* 94: 263–78.

——(1995), *Naturalizing the Mind* (Cambridge, Mass.: MIT Press).

Ericsson, K. A., and Simon, H. A. (1993), *Protocol Analysis: Verbal Reports as Data* (rev. edn., Cambridge, Mass.: MIT Press).

Evans, G. (1982), *The Varieties of Reference* (Oxford: Oxford University Press).

Ganis, G., Thompson, W. L., and Kosslyn, S. M. (2004), "Brain Areas Underlying Visual Mental Imagery and Visual Perception: An fMRI Study", *Cognitive Brain Research* 20: 226–41.

Gettier, E. (1963), "Is Justified True Belief Knowledge?", *Analysis* 23: 121–3.

Goldenberg, G., Müllbacher, W., and Nowak, A. (1995), "Imagery Without Perception—a Case Study of Anosognosia for Cortical Blindness", *Neuropsychologia* 33: 1373–82.

Gordon, R. M. (1996), "'Radical' Simulationism", in P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind* (Cambridge: Cambridge University Press).

Haber, R. N. (1979), "Twenty Years of Haunting Eidetic Imagery: Where's the Ghost?", *Behavioral and Brain Sciences* 2: 583–629.

Hume, D. (1740/1978), *A Treatise of Human Nature*, ed. L. A. Selby-Bigge (Oxford: Oxford University Press).

Hurlburt, R. T. (1990), *Sampling Normal and Schizophrenic Inner Experience* (New York: Plenum Press).

——(1993), *Sampling Inner Experience in Disturbed Affect* (New York: Plenum Press).

——(2009), "Unsymbolized Thinking, Sensory Awareness, and Mindreading", *Behavioral and Brain Sciences* 32: 149–50.

Hurlburt, R. T., and Akhter, S. A. (2008), "Unsymbolized Thinking", *Consciousness and Cognition* 17: 1364–74.

Intraub, H., and Hoffman, J. E. (1992), "Reading and Visual Memory: Remembering Scenes That Were Never Seen", *American Journal of Psychology* 105: 101–14.

Jones, S. R., and Fernyhough, C. (2007), "Thought as Action: Inner Speech, Self-Monitoring, and Auditory Verbal Hallucinations", *Consciousness and Cognition* 16: 391–9.

Kosslyn, S. M. (1994), *Image and Brain: The Resolution of the Imagery Debate* (Cambridge, Mass.: MIT Press).

Kosslyn, S. M., Thompson, W. L., and Ganis, G. (2006), *The Case for Mental Imagery* (Oxford: Oxford University Press).

Langland-Hassan, P. (2008), "Fractured Phenomenologies: Thought Insertion, Inner Speech, and the Puzzle of Extraneity", *Mind and Language* 23: 369–401.

Lycan, W. G. (1987), *Consciousness* (Cambridge, Mass.: MIT Press).

——(1996), *Consciousness and Experience* (Cambridge, Mass.: MIT Press).

Mackay, D. G. (1992), "Constraints on Theories of Inner Speech", in D. Reisberg (ed.), *Auditory Imagery* (Hillsdale, NJ: Lawrence Erlbaum).

McKinsey, M. (1991). "Anti-Individualism and Privileged Access", *Analysis* 51: 9–16.

Moran, R. (2001), *Authority and Estrangement* (Princeton: Princeton University Press).

Nichols, S., and Stich, S. (2003), *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Oxford: Oxford University Press).

Ryle, G. (1949), *The Concept of Mind* (London: Hutchinson, Peregrine Books).

Shoemaker, S. (1988), "On Knowing One's Own Mind", *Philosophical Perspectives* 2: 183–209.

——(1994), "Self-Knowledge and 'inner sense'", *Philosophy and Phenomenological Research* 54: 249–314. Page reference to the reprint in Shoemaker (1996).

——(1996), *The First-Person Perspective and Other Essays* (Cambridge: Cambridge University Press).

Smith, A. D. (2001), "Perception and Belief", *Philosophy and Phenomenological Research* 62: 283–309.

Sosa, E. (1999), "How to Defeat Opposition to Moore", *Philosophical Perspectives* 13: 141–53.

Stone, T., and Young, A. (1997), "Delusions and Brain Injury: The Philosophy and Psychology of Belief", *Mind and Language* 12: 327–64.

Thomasson, A. (2003), "Introspection and Phenomenological Method", *Phenomenology and the Cognitive Sciences* 2: 239–54.

Williamson, T. (2000), *Knowledge and Its Limits* (Oxford: Oxford University Press).

Zatorre, R. J., and Halpern, A. R. (2005), "Mental Concerts: Musical Imagery and Auditory Cortex", *Neuron* 47: 9–12.