

Beliefs-in-a-Vat

I Introduction

The over-arching claim that I intend to defend in this paper is that while widespread ‘local’ error is conceivable, we cannot, in the end, make sense of the radical sceptical idea that no one has ever been in touch with an ‘external world’ – that perception might not even in principle be a capacity for knowledge. To this end, I will show that an asymmetry exists between ‘local’ and ‘global’ sceptical scenarios, such that the possibility of ‘local’ error does not imply that ‘global’ error must also be possible. Instead, we will see that what gives rise to the radical sceptical problem is an unquestioned acceptance of the ‘New Evil Genius Thesis’ (NET) – the notion that I and my ‘envatted’ counterpart share the same perceptual experiences, even though my benighted twin has never had any contact with an ‘external’ reality. Although most contemporary epistemologists take NET for granted, I will show that it cannot, ultimately, be rendered intelligible, and, consequently, that neither can the ‘global’ sceptical scenario that depends on it.

The strategy pursued in the paper is as follows. In the next section I argue that ‘local’ and ‘global’ sceptical scenarios are significantly different, and that arguments from perceptual illusion actually undermine, rather than provide good reasons for, radical scepticism. In section III I show that the ‘global’ sceptical scenario assumes the truth of NET, which, in turn, presupposes that meaning and belief-ascription are possible in contexts where all I can have access to is my ‘internal’ paraphernalia. I argue that we cannot, in the end, make sense of such a conception, and that it is consequently impossible to maintain with the

sceptic that knowledge of the content of our perceptual beliefs can be preserved, while knowledge of an ‘external world’ is jettisoned. For if content ascription is impossible in the radical sceptical scenario, then no ‘global’ scepticism can even be formulated: without the availability of the thought that the world at least *appears* to us in a certain way, there can be no sceptical concern about the veridicality of these appearances (or, indeed, about anything). In other words, it is not that radical scepticism is false. Rather, if I am right, the ‘global’ sceptical scenario turns out, in the end, to lack coherence¹.

II ‘Local’ Sceptical Scenarios and Arguments from Perceptual Illusion

One of the standard ways of motivating radical scepticism consists of trying to show that if it is conceivable for something sometimes to be the case, it is also possible to imagine that it could *always* be the case. Or, more pertinently, if one can sometimes believe that P even though P is not the case, it is possible that one could *always* believe that P even though P is not the case. I will call this type of argument an ‘aggregate argument’, or the attempt to get to a ‘global’ or ‘radical’ sceptical context by way of ‘aggregating’ cases of ‘local’ error. I will show that such ‘aggregation’ attempts fail, and, hence, that ‘global’ sceptical scenarios cannot be constructed out of ‘local’ ones.

That ‘aggregate arguments’ appear, at first blush, intuitive probably stems from the fact that insufficient attention tends to be paid, in contemporary epistemology, to the differences between ‘local’ sceptical cases on the one hand – the possibility of my being wrong about many of my beliefs – and ‘global’ or ‘radical’ sceptical scenarios on the other –

¹ If radical scepticism were false, it would specify an intelligible scenario that happens not to obtain. What I am attempting to show, however, is that radical scepticism is an illusion, not a falsehood.

the thought that all our perceptual beliefs could be false, since no one might *ever* be in touch with an ‘external world’² at all. When sceptical ‘brain-in-a-vat’ (BIV) scenarios are considered in epistemology, it is frequently left unclear what the ‘scope’ of such cases is supposed to be. Are they meant, for instance, to attack the possibility of ‘local’ knowledge-possession, mine say – e.g. if I suddenly became ‘envatted’ but was previously ‘normal’; or if only *I* were envatted, but everyone else were ‘normal’ – or are they supposed to undermine the possibility of anyone’s ever possessing (or having possessed) knowledge of anything (the ‘global’ or ‘radical’ sceptical scenario)? Since the two cases are distinct, it is not sufficient to appeal to ‘local’ fallibility, run an ‘aggregate argument’, and obtain the conclusion that, for all we know, we might be in the ‘global’ situation. For, as we shall see below, the fact that we can sometimes be wrong does not entail, but rather *precludes*, that we could *always* be wrong.

So, if we take the BIV scenario, not as a metaphor, but literally, i.e. as actually specifying a scenario where I have become envatted and all my current impressions are being generated by electrodes, then this will be a description of a ‘local’ sceptical case with certain determinate implications: that there is a world, for example, in which the vat containing my brain exists; that there are evil scientists (or robots or aliens or what-have-you) who have caused me to become envatted (or who have ‘bred’ ‘me’ – my brain – to be envatted); that it is scientifically possible to separate brains from bodies without killing off the brain; that I might find out about my previous envatted state by, for instance, having my brain ‘reinserted’ into my body; that I ‘am’ my brain and so on. This, as it were, literal fleshing-out of the sceptical scenario is precisely what turns it into a ‘local’ case: if I imagine that I might be the victim of such a predicament, this does not imply that anyone else is, or, indeed, that the

² I have put quotation marks around ‘external world’, as if I am right that we cannot, in the end, make sense of a global scepticism, then neither can we make sense of the thought that we could be locked into an ‘internal’ world of appearances, which could be as they are, even though there be nothing ‘outside’ of us.

‘external world’ as such does not exist (since, as we have just seen, a world containing vats, brains and evil scientists is in fact presupposed by this form of scepticism). All it implies is that, for as long as I am a BIV, most of what I believe about the world is false³.

The foregoing parallels what so-called ‘arguments from illusion’ can achieve in epistemology, since here, too, the possibility of local fallibility does not imply the possibility of global error. That is to say, just as the ‘local’ does not imply the ‘global’ sceptical scenario, so the possibility of perceptual error or illusion does not imply systematic or ‘global’ perceptual unreliability. For example, I am only able to determine that when I look at a square tower from a distance, it will appear round, because I can trust my perception that from close-by it looks square, and there is an explanation available that can tell me why it nevertheless appears round from some way off. If perception in general were deceptive, I could not make the judgement that perceptual appearances are sometimes misleading. All I could do would be to report, for instance, that at time t_1 I have the impression that ‘thing 1’ that I see is square, while at time t_2 I have the impression that ‘thing 2’ that I see is round. And since ‘thing 1’ and ‘thing 2’ might, for all I know, be different things, I could not even conclude that one perceptual experience might be an accurate representation of the way things are, while the other might not. But if I cannot make this judgement, I am similarly unable to conclude that at time t_2 I am being misled, for my perceptual experience at t_2 would only be misleading if it were an experience of the same thing that I encountered at t_1 , so that my reports at t_1 and t_2 would turn out to be in conflict with each other. As long as I have no reason for assuming that my perceptual experiences of ‘thing 1’ and ‘thing 2’ are in fact experiences of the same physical object, however – which, of course, I would not if

³ This distinction matters, because in order to get such a ‘local’ BIV scenario off the ground, the sceptic needs to make the previously specified assumptions about the world, which implies that, generally speaking, knowledge of such a world must, in principle, be possible (even if I can have no access to this knowledge while I am a BIV). I contrast this with a ‘radical’ or ‘global’ sceptical scenario which I take to attack the very idea that knowledge of the world is *ever* possible (I call this the ‘metaphorical BIV scenario’. See discussion below).

perception were generally defective (for then I would have no grounds for trusting one report more than the other; I would rather have to assume they are equally misleading) – there is no way of determining that a perceptual ‘illusion’ has in fact occurred. All I could say is that I am having different perceptions at different times, but this, of course, does not suffice to allow me to infer that at time t_2 I was misled, and, hence, that perceptual errors are possible⁴. Consequently, far from showing that perception, in general, is defective, the possibility of perceptual error actually presupposes that perception is generally in good working order. Hence, ‘aggregate arguments’ based on the possibility of perceptual error fail: I am not entitled to infer that because I can sometimes be wrong (‘local case’), it is possible that I could always be wrong (‘global case’).

In other words, just as we need to presuppose the existence of an ‘external world’ that is broadly like our own in order to get the ‘literal’ BIV scenario off the ground, so perception must generally be taken to be reliable if an ‘argument from illusion’ is to be constructed.⁵ So, what makes ‘local’ sceptical scenarios possible is the very thing that they are, ironically, drafted in to undermine: a background of generally veridical perceptual experience. In this

⁴ What if we suppose that, at t_1 , I have the impression that thing 1 that I see is square, at t_2 I have the impression that thing 2 that I see is round, and between t_1 and t_2 I have the impression that thing 1 hasn’t changed shape or been replaced – wouldn’t we then have to conclude that a perceptual illusion has in fact occurred? We might indeed conclude in such a case that a perceptual illusion has occurred (as we might also in the other scenarios), but the relevant point is that we can only draw this conclusion because we are taking our impression that thing 1 has not changed between t_1 and t_2 (or been replaced) to be veridical, which confirms what I was trying to show: namely, that we cannot conclude that a perceptual error has occurred unless we are willing to grant that some of our sense-perceptions can be veridical. If we don’t, then we certainly cannot grant that we can take our impression that thing 1 hasn’t changed between t_1 and t_2 at face value (and hence we could not conclude that a perceptual error has occurred). In other words, conflicting impressions by themselves are insufficient to motivate the thought that at least one of these impressions must be false unless we have already accepted that these impressions are impressions of ‘externally’ existing things that have a stable nature. If impressions are not taken to convey information about externally existing (stable) things, there is nothing about the impressions themselves that would give us reason to think that we can’t have different ones at different times.

⁵ This has nothing to do with the question of whether or not ‘local’ error is always detectable. It may not be detectable in any given case. The point is rather that if I start with the assumption that *all* perceptions are unreliable, then I can *never* formulate an argument from illusion, as such an argument presupposes that some of our perceptions are reliable (e.g. that the tower has not changed shape and now looks square). Hence, I cannot use an argument from illusion in order to undermine the very thing the argument itself presupposes: the reliability of some of our sense-perceptions. So I am not endorsing the false principle that if it is not possible to tell that p , it is not possible that p .

respect, the ‘literal’ BIV scenario is just a souped-up, ‘sci-fi’ version of the more pedestrian arguments from illusion and cannot, for this reason, achieve more than they can: to show that perceptual error is possible (that human beings are fallible). Since we have already seen, however, that ‘aggregate arguments’ are fallacious, an appeal to perceptual fallibility alone is not enough to get one the conclusion that it is possible that no one might ever have had true perceptual beliefs about the world at all. If sceptical scenarios, therefore, amounted to no more than the local variety, one wouldn’t have much reason to be concerned about them⁶.

III Content in the ‘Global’ ‘Vat-World’

III.1 The BIV Hypothesis as Metaphor

In its more virulent form, however, the BIV hypothesis operates as a metaphor: i.e. it is not meant to be taken literally (as described above), but as an expression of the rather more serious worry that even in the best possible case, perception is never factive or able to engage with an ‘external world’. It is, in other words, emblematic of the concern that there might be a systematic mismatch between everything we collectively believe to be the case about the world, and the way the world actually is. And, if this were in fact so, then, even if it seemed

⁶ One might, perhaps, wish to object here that even if the foregoing is correct, it is at least *logically* possible that I am currently envatted and only vainly imagining that I am writing a paper (or, to speak with Descartes, that I am currently dreaming that I am doing this). Although this is true, it is important to bear in mind that the mere fact that I am able to conceive of such a state of ‘envatment’ does not give me a real reason to suppose that I might *actually* be the victim of such a situation. Why not? Because, in the absence of a general argument designed to undermine the possibility of perceptual knowledge *per se* (not just my own), I have no real grounds for thinking that the imagined scenario might be the actual one. E.g., the state of current science is not yet advanced enough to make such scenarios even empirically possible; there is no evidence of alien or robot activity etc. That is to say, the onus is not on the anti-sceptic to rule out every conceivable logical possibility in advance, unless the sceptic can give one a reason for supposing that knowledge of the ‘external world’ is never possible (and, if I am right, that’s what – so far – he has not done).

otherwise, we would never have conclusive perceptual knowledge of anything; nor could we ever find out (even in principle) whether such a scenario obtained, as any form of evidence one could appeal to would itself be part of the ‘grand illusion’.

Now what is it that so much as gives us the idea that all of our beliefs taken together could fail to ‘match’ the way the world is given that, as we have just seen, we are not entitled to infer this from ‘local’ sceptical scenarios? The underlying reason, which tends to manifest itself in a variety of different ways, seems to be the thought that we are somehow not in direct touch with the objects that make up the physical world, but only with our sense-experiences or mental states from which one must try to work out ‘what is going on out there in the external world’ (White (2014: 299)). Proponents of such a notion⁷ therefore take it for granted that it is possible to know how things subjectively seem to one, even though it is conceivable that one’s perceptual experiences *as a class* may never be experiences *of* anything. Hence, on this conception, ‘appearances’ must be conceived as having a determinate content which one can be presented with in sense-experience – for example, an appearance ‘as of there being a cookie in front of one’ as opposed to an appearance ‘as of there being a computer on the table’ – even though there might be no physical objects (such as cookies and computers) at all. Such a Cartesian picture of one’s epistemological situation presupposes that the content of one’s experiences is completely detachable from what these experiences are putatively experiences of, and, consequently, that one can know the content that one is presented with in sense-experience – say ‘seeing books on the table’ – without thereby necessarily coming to know anything at all about what these experiences reveal about the physical world itself (including if there is one).

It seems clear that such a conception of experience is the ‘intuition’ behind NET: the thought that the experiences that I share with my ‘envatted’ counterpart, and that I base my

⁷ I call this the ‘Cartesian picture’ in my *The Illusion of Doubt* (2016). It is the target of the book.

perceptual knowledge on, are, in all relevant ‘internal’ respects, the same. For if this thesis is coherent, it entails that one’s putative knowledge-claims about the world can be derived solely from knowledge of how things appear to one, since unless one can rule out that one is in the ‘global’ ‘vat-world’, one can never have knowledge of anything else. Given that it must be possible to express these alleged ‘knowledge-claims’ in a language (if they are to be expressed at all), this conception further implies that a language must be possible that does not presuppose any acquaintance with an ‘external world’, or with ‘other minds’, as if we were in the ‘global’ ‘vat-world’, there would be no such things. Consequently, the words of this language would have to derive their meaning purely from being linked to episodes of one’s own consciousness (since there may be nothing else), which means that it must be possible to identify and describe these episodes independently of whether they ever inform one of ‘reality’ or of ‘the facts’. Should it turn out, therefore, that this is impossible – that, in effect, there can be no such thing as meaning and belief ascription in the ‘global’ sceptical scenario – then this also shows that the radical sceptic cannot have it both ways here: that she cannot hang on to the notion that knowledge of the content of one’s sense-experiences is possible, while at the same time maintaining that knowledge of the ‘external world’ is forever foreclosed⁸.

III.2 Belief and Interpretation

Now Davidson has famously argued that it is impossible to interpret someone with preponderately false beliefs, as beliefs can be ascribed only if they are, in the main, veridical. If Davidson is right about this, and it is a condition of having an interpretable language that

⁸ And without meaning and belief-ascription, no ‘global’ sceptical scenario can even be formulated.

most of what a speaker says is true, then, given that we *are* able to understand and interpret others⁹, it is incoherent to suppose that radical perceptual error is possible.

Given the far-reaching consequences for radical scepticism of Davidson's argument, it is worth taking a closer look at the details. In 'A Coherence Theory of Truth and Knowledge', Davidson contends that there is an intimate connection between theories of meaning and epistemology, as we need an answer to the question of how one determines that a sentence is true (Davidson 1986: 312). This question raises the same kind of difficulties as the corresponding question in epistemology about what justifies our beliefs. Two responses, according to Davidson, are possible: one is coherentist – seeking the justification of beliefs only in other beliefs – and the other is foundationalist – attempting to anchor at least some sentences held true to 'non-verbal rocks' (*ibid.*)¹⁰. Davidson suggests that we should give up the idea that meaning or knowledge can be grounded on something that counts as an ultimate source of justification (Davidson 1986: 313), as the search for an empirical foundation (i.e. 'intermediaries' of some kind, such as 'sense-data', 'experience', the 'Given') merely leads to scepticism:

For if the intermediaries are merely causes, they don't justify the beliefs they cause, while if they deliver information, they may be lying. Since we can't swear intermediaries to truthfulness, we should allow no intermediaries between our beliefs and their objects in the world. Of course there are causal intermediaries. What we must guard against are epistemic intermediaries' (Davidson 1986: 312).

⁹ Given footnote 8, the sceptic should grant that we are interpretable. Nevertheless, I will, in section III.4, consider the objection that there might be no meaning.

¹⁰ One might not wish to accept how Davidson carves up the epistemological space here – there may be more options than just foundationalism or coherentism. But even if one accepts neither, this makes (on my reading, at least) no difference to Davidson's overall argument. For more on these issues, see my (2016).

In other words, if experience is conceived as merely a ‘brute’ cause in a radically externalist sense, then it cannot enter into the ‘space of reasons’¹¹ and provide justification for our perceptual beliefs, while if it is thought of in roughly Cartesian terms as consisting of an ‘internal realm’ of fully conceptualized private intermediary items which may or may not accurately ‘represent’ the ‘external world’, it leads to scepticism¹². Since Davidson rightly regards this dilemma as fatal, the only plausible alternative is to cut out the middle-man and to connect our perceptual beliefs not with ‘experiential intermediaries’, but with the objects in the world that cause our beliefs: ‘In the plainest and methodologically most basic cases’, we must take ‘the objects of a belief to be the causes of that belief’ (Davidson 1986: 317-8)^{13 14}.

But how, in the context of ‘radical interpretation’, do we know which objects are causing someone’s beliefs, given that we don’t yet know what their sentences mean? It is here that Davidson’s ‘principle of charity’ comes in – directing the interpreter to translate so as to read some of his own standards of truth into the pattern of sentences held true by the speaker (Davidson 1986: 316). So, for example, if the interpretee utters the word ‘rabbit’ whenever a rabbit is scampering past, then it is reasonable to assume, *ceteris paribus*, that ‘rabbit’ means *rabbit*. Of course, we can sometimes be wrong about this, but if someone is supposed to be interpretable, we cannot start with the assumption that *all* of his beliefs about the world might be false. That is to say, it cannot be *the rule* that interpreter and interpretee

¹¹ See McDowell (1994).

¹² I am of course not taking this as a reason that counts against *scepticism*, but as a *prima facie* reason that speaks against the Cartesian conception. The conception itself will be attacked directly in the next section.

¹³ Is that just to assume content externalism? No, for as we will see below, endorsing the Cartesian conception requires presupposing the Myth of the Given, and this is hopeless. Davidson attacks the Myth directly in his (2001).

¹⁴ In other words, the target of a linguistic response is its normal cause. As McDowell (2003) points out, however, this thought risks being misleading if it suggests that an interpreter must first identify ‘the’ cause of a linguistic response and then, on the basis of having independently settled such questions, declare that cause to be the target. Rather, identifying ‘the’ cause is itself already an exercise of interpretation (McDowell 2003: 678-9). I think that McDowell is right about this, but don’t have the space to go into this issue here (it is discussed in my (2016) in respect to Wittgenstein’s rule-following considerations).

understand each other on the basis of shared but completely erroneous beliefs. Why not? Because we don't first form concepts, and then 'discover' what they apply to. Rather, we learn both at the same time: 'light dawns gradually over the whole'¹⁵. Consequently, it is idle to think that we could fall into massive error by accepting this principle, for it is non-optional: 'We make maximum sense of the words and thoughts of others when we interpret them in a way that optimizes agreement' (Davidson 2001: 197).

What are the implications of this for radical scepticism? If Davidson's account is correct, it seems that we cannot make sense of a BIV's 'language', as *ex hypothesi* all the perceptual beliefs of a BIV are supposed to be false. If Davidson is right, however, that it is a condition of content ascription that some of the BIV's beliefs must, on pain of uninterpretability, be veridical, then these beliefs are either not about the sorts of things that would generate the sceptical worry (i.e. if veridical, they would have to be interpreted as being about electrode stimulations, not about, say, tables and trees), or it is impossible to assign beliefs to the BIV altogether. Either way, Davidson's argument undermines NET: the BIV's perceptual beliefs are either different from those of its non-envatted twin, or it is impossible to ascribe coherent beliefs to the BIV at all.

That Davidson's argument seems to have these anti-sceptical consequences is, for Colin McGinn, a *prima facie* reason to think that it must be false. For how can semantic considerations show that radical scepticism is mistaken? *Pace* Davidson, McGinn wants to maintain that we *can* make sense of experience *qua* 'epistemic intermediary', and that consequently, *contra* Davidson and Putnam, some concepts *are* in the head: 'Experience can ground concept possession if it can represent the property denoted by the concept in question but it cannot do so if the property is not *capable* of being represented in experience (i.e., if it is not a property relating to the *appearance* of things)' (McGinn 1986: 362). In other words,

¹⁵ Ludwig Wittgenstein, *On Certainty* §141.

McGinn believes that a half-way house is possible here: on his view, observational concepts (such as ‘red’ and ‘round’, for instance) are ‘in the head’, while non-observational concepts are dependent on ‘external’ factors. This means that, on McGinn’s conception, it is possible to attribute observational concepts to a BIV, even though the beliefs these concepts figure in could all be false. So, on McGinn’s ‘hybrid’ view, the sceptical problem re-emerges (at least for the observational concepts). In the remainder of this section, I will argue that McGinn’s contention that ‘experience’ can ground concept possession if the experience can ‘represent’ the property in question is misconceived and consequently does not pose a challenge to Davidson’s view.¹⁶

III.3 Can a ‘BIV from Birth’ Have a Language?

Recall that if you are in the ‘global’ sceptical scenario, then you are, as it were, a ‘BIV from birth’ who has never interacted with other physical beings and objects at all, and whatever you take to be an experience of a physical object may be nothing more than an episode of your own consciousness ‘masquerading’ as an experience of something ‘external’. But, if so, this implies that you have been able to derive all your perceptual concepts from your own private experiences (the episodes of your consciousness), since in the ‘global’ ‘vat-world’ there is nothing else. This raises the question: is such a ‘private experience language’ possible?

Let’s take the putative experience of ‘red’ – if you are in the ‘global’ ‘vat-world’ and there are no real red objects for you independently to refer to, how can you know what kind

¹⁶ Even if McGinn’s ‘hybrid’ view is mistaken, someone might object to Davidson’s argument by contending that *all* concepts are ‘in the head’. But, if I am right, the argument below also rules out the intelligibility of *that* notion.

of experience you are having? For unless you presuppose a version of what Sellars (1997) calls the ‘Myth of the Given’ – that an experience of ‘red’ is, as it were, a ‘readymade’ internal introspectible item that just ‘tells you’ that it is an experience of red¹⁷ – how can you differentiate a ‘red’ experience from all your other mental furniture? Given that you don’t have anything independent and ‘external’ to appeal to, but only episodes of consciousness that stand on the same level as the experience you are currently trying to identify, it seems that unless you question-beggingly start with the assumption that your experiences are somehow ‘self-identifying’, it is difficult to see how ‘naming’ and identifying these private experiences is supposed to work.

In other words, and as Wittgenstein has famously pointed out, one cannot just take it for granted that one can give a sign a meaning by, say, mentally uttering a sound in the presence of a certain (‘internal’) phenomenon. Why not? Because this ‘sound uttering’ must be able to provide one with rules for the correct use of the word and these cannot just be ‘read off’ from the episode of consciousness itself. For unless one already has prior knowledge of the ‘post where the new word is to be stationed’ (*Philosophical Investigations*¹⁸ §257), one will not understand the ostensive definition, ‘this is called “tove”’¹⁹, uttered in the presence of a pencil, for example, either. Given that ‘tove’ might mean a whole host of different things – for instance, ‘writing utensil’, ‘number’, ‘physical object in general’, ‘position on the table’, ‘colour’, ‘sharp’, ‘blunt’ etc. – unless one already knows that someone means ostensively to define the *name of a particular writing utensil*, say (which in turn presupposes that one has already acquired the concept ‘writing utensil’), one is not going to understand the ostensive definition, as one will have no idea what the ‘this’ in ‘this is called “tove”’ is supposed to

¹⁷ Effectively the Cartesian view.

¹⁸ PI henceforth.

¹⁹ See Ludwig Wittgenstein, *The Blue Book*, p.2.

refer to. In short, ostensive definition underdetermines the *definiendum* unless some prior linguistic competence is already present.

But things are even worse in the ‘BIV from birth’s’ case. For the ‘BIV from birth’ (BFB henceforth) could only give itself a *private* ostensive definition, as *ex hypothesi* it has no access to anything public and external (such as physical pencils):

Let’s imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign ‘S’ and write this sign in a calendar for every day on which I have the sensation. – I first want to observe that a definition of the sign cannot be formulated. – But all the same, I can give one to myself as a kind of ostensive definition! – How? Can I point to the sensation? – Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation – and so, as it were, point to it inwardly. – But what is this ceremony for? For that is all it seems to be! A definition serves to lay down the meaning of a sign, doesn’t it? – Well, that is done precisely by concentrating my attention; for in this way I commit to memory the connection between the sign and the sensation. – But “I commit it to memory” can only mean: this process brings it about that I remember the connection *correctly* in the future. But in the present case, I have no criterion of correctness. One would like to say: whatever is going to seem correct to me is correct. And that only means that here we can’t talk about “correct” (PI §258).

Given how compressed Wittgenstein’s argument is, it is usually thought²⁰ that its aim is to attack the reliability of memory in the private context, when its actual target is the very idea of private ostensive definition as such, something that has nothing to do with a general scepticism about memory. Philosophers who endorse the ‘memory sceptical’ reading believe that Wittgenstein is claiming that there is no such thing as ‘private’ rule-following, since in the private scenario no distinction can be drawn between what seems right to me, and what is *actually* right. They then go on to link this idea to the thought that, in the private case, where

²⁰ See for example, A.J. Ayer (1971), P.F. Strawson (1971), Helen Hervey (1971), C.W.K. Mundle (1971), Judith Jarvis Thomson (1971), Saul Kripke (1982), Robert Fogelin (1987).

I cannot appeal to the judgement of others, I have no way of checking whether my present sensation, which I take to be ‘S’ again, actually corresponds to the sample I originally concentrated my attention on and thus labelled ‘S’ in the first place. In other words, it may be that I misremember which sensation is supposed to be ‘S’, and, since I have nothing outside of myself to appeal to – as it were no ‘external checks’ to corroborate what I believe to be the case – whatever is going to seem right to me, is right, and that just means that one can’t talk about right (PI §258).

The problem with this reading is not only that it undermines the possibility of a public language just as much as of a ‘private’ one²¹ (and is, for this reason, not likely to be correct²²), it also already concedes the main point to the sceptic or ‘private linguist’. For Wittgenstein is not attacking the notion that I will not be able correctly to identify *future* instances of ‘S’; instead, he is trying to undermine the thought that *any* connection has so far been set up between the putative internal episode and the sign ‘S’. After all, the question of future correct identification presupposes that the sign ‘S’ has already been given a meaning, and that is just what is at issue.

In other words, ‘memory sceptical’ readings already start with the assumption that meaning is possible in the ‘global’ ‘vat-world’, since they grant to the BFB that it has been able to set up a connection between ‘S’ and the ‘internal episode’ when the coherence of that notion in the radical sceptical scenario is the very thing that is in question. For if I assume that I already have the concept of a sensation, then, naturally, I can introduce new words for as-yet-unheard-of sensations. But given that ‘sensation’ is itself a public concept, the private linguist cannot just help himself to this notion without showing how it can be acquired purely ‘internally’ – i.e. without presupposing a connection to anything ‘outer’ (such as pain

²¹ As public language, too, relies on our being able to remember the rules for the correct use of words.

²² For an extensive discussion of the exegetical situation, see my (2016).

behaviour, daggers etc.), or other public concepts (such as ‘expressing pain’, say). Rather, the private linguist must be able to derive the concept solely by ‘concentrating his attention’ on a private, internal phenomenon, as, *ex hypothesi* this is all that he has access to. But, if so, then an appeal to a ‘memory of S’ won’t help him, since to have a ‘memory of S’ presupposes that he already knows what ‘S’ means:

Let us imagine a table, something like a dictionary, that exists only in our imagination. A dictionary can be used to justify the translation of a word X by a word Y. But are we also to call it a justification if such a table is to be looked up only in the imagination? – ‘Well, yes; then it’s a subjective justification.’ – But justification consists in appealing to an independent authority – ‘But surely I can appeal from one memory to another. For example, I don’t know if I have remembered the time of departure of a train correctly, and to check it I call to mind how a page of the timetable looked. Isn’t this the same sort of case?’ No; for this procedure must now actually call forth the *correct* memory. If the mental image of the timetable could not itself be *tested* for correctness, how could it confirm the correctness of the first memory? (As if someone were to buy several copies of today’s morning paper to assure himself that what it said was true) (PI §265).

Far from making any kind of verificationist point here, what Wittgenstein is really doing is accusing the private linguist of begging the question. For to call up a memory in order to confirm something, presupposes that the memory in question is indeed a memory of the sort of thing that will, if the memory is accurate, confirm the thing in question. That is to say, it is only if my ‘memory’ is a memory of a *train timetable* (and is accurate) that it will confirm the time of departure of the train. If I called up the memory of a page in my Gordon Ramsey recipe book instead, say, then the memory, even if accurate, would not confirm the train departure time. In other words, the private linguist is only entitled to appeal to his ‘memory of S’ as a way of confirming that his present internal episode is S, if the ‘memory of S’ is a

memory of *S* (and not, say of ‘T’, ‘U’ or ‘XYZ’)²³. But, again (as in PI §258), the private linguist can only know that it is a memory of *S* if he already knows what ‘S’ means. It is the appeal to the ‘memory of S’, however, that is supposed to provide the criterion for what ‘S’ means, so the private linguist is, in effect, using the memory of the meaning of *S* to confirm itself. It is for this reason that Wittgenstein says that what the private linguist is doing is like buying several copies of the (same) morning paper in order to assure himself that what it says is true. If this is right, then *pace* ‘memory sceptical’ readings, it is actually irrelevant whether or not one’s memory deceives one, since if one doesn’t yet know what ‘S’ means, there is nothing for one’s memory to deceive one about.

What Wittgenstein’s ‘anti-private language’ dialectic is, therefore, supposed to show is that in the ‘global’ ‘vat-world’, ‘S’ will remain a meaningless sign, since introspection alone cannot give it a use. That is to say, Wittgenstein is trying to undermine the old myth that we can somehow ‘read off’ from the ‘object’ itself the way that its ‘name’ is to be applied²⁴ – a myth that is helped on its way by what Putnam calls a ‘pictorial semantics’ (Putnam (2002: 15)) – the empiricist notion that words refer to ideas which are mental ‘copies’ of the objects we perceive. Even if none of these notions are explicitly endorsed by most contemporary philosophers, they form part of the inherited (and, perhaps, ‘intuitive-seeming’) background that gives sustenance to the idea that a ‘private experience language’ must be possible. For with this picture in place, it is going to seem natural to think that just as we learn what ‘cow’ means by looking at cows, so we might learn what ‘sensation’ means by, as it were, ‘looking at’ our sensations. But is it sufficient to give a word a meaning that it be uttered when and only when a cow is present? No. Malcolm makes the point well:

²³ Compare Kenny (2006: 152).

²⁴ This idea is similar to what Putnam (1981:51) calls a ‘Magical Theory of Reference’ – the view that some occult ‘noetic’ rays connect a word with its referent; David Wiggins (1980) calls this notion a belief in ‘Self-Identifying Objects’.

The sound might refer to anything or nothing.²⁵ What is necessary is that it should play a part in various activities, in calling, fetching, counting cows, distinguishing cows from other things and pictures of cows from pictures of other things. If the sound has no fixed place in activities ('language-games') of this sort, then it isn't a word for *cow*. To be sure, I can sit in my chair and talk about cows and not be engaged in any of those activities – but what makes my words *refer* to cows is the fact that I have already mastered those activities; they lie in the background (Malcolm (1971:35)).

In other words, just as human agreement in judgements is a necessary 'background condition' for the possibility of rule-following (or belief-ascription), so the role that a word plays in various 'language-games' provides the necessary 'background' that gives it a meaning. But, if so, then it is not possible to learn a word's meaning just by 'looking at' the thing it is supposed to denote, as doing so will not give one insight into how this word is actually used in the various practices in which it is at home. Consequently, a mistaken conception of how we acquire concepts in the ordinary, public sphere, when applied to the 'internal realm' conceived as a kind of inner analogue of an 'outer' 'external' world, is naturally going to suggest to us that a 'private experience language' must be possible.

In short, unless we plump for the 'Myth of the Given' and a Magical Theory of Reference (see footnote 24), it is hard to see how the BFB can get a 'private experience language' off the ground. For if I have nothing outside of my own consciousness to appeal to, how, except magically, can I learn to discriminate between these highly volatile and transient 'internal' goings-on? Without a connection to anything 'outer' and 'external', it's difficult to conceive how the BFB can so much as identify its episode of consciousness as being of a certain kind. And if the BFB cannot pick out any stable particular, then it cannot try and 'name' that 'thing' either. For, as we have already seen, in order to understand an ostensive

²⁵ Compare Boghossian (2002).

definition, I must already know what specific feature of the object in question is being singled out for naming. But this seems impossible, if there is nothing independent and external that I can appeal to as a reference point. It is here that Wittgenstein's and Davidson's arguments converge.

If this is right, however, then McGinn's objection to Davidson misfires: there is no half-way house that, as it were, allows epistemic intermediaries in through the backdoor. For, if, for example, I do not learn what 'red' (or any other observational concept) means by being presented with a self-subsistent, 'internal' 'intermediary item' (an unconceptualized Given) thought to form part of an 'interface' of 'experiences' which 'intervenes' between myself and the 'external world', and from which one can somehow 'magically' 'read off' how the word 'red' is to be used, then there is no longer any room either for the thought that what lies beyond the 'interface' might be radically different from what we think it is.

That is to say, if we all, collectively, were BFBs (the radical sceptical scenario²⁶), then the 'experiences' that we would allegedly be being 'fed' could only have the *same* content as our real, 'external world' experiences, if it is taken for granted that the putatively phenomenologically indistinguishable BIV-experiences possessed some kind of 'intrinsic' content that could be 'accessed' merely by 'inspecting' these 'inner experiences' themselves. Since such 'private inspection' could provide us with the relevant concepts only if private ostensive definition (and 'magical' acts of reference) were possible, the arguments advanced against this notion show that there is, in the end, no making sense of the idea that we could be wrong about everything all of the time.

²⁶ Of course, even speaking of 'BIVs from birth' is metaphorical in the sense that one can only get a 'global' sceptical scenario out of such a conception if one disregards that there has to be something 'outside' of these vats, and furthermore someone (or something) to 'service' them.

III.4 Some Objections

But perhaps this was too quick. For what if someone argued that McGinn's objection just doesn't go far enough inasmuch as we *all* need to derive our concepts purely from our 'own experience' and not from anything 'outer' whose presence we can at best infer? Here is one such suggestion:

Wittgensteinians who understand learning a word 'from one's own case' in this way [i.e. from one's own experience] tie the theorists they oppose to an innocent, unavoidable practice. 'There are words a person can learn only through his own (sensory) experience' is an innocent truth. Even a non-sceptic, who entertains no doubts about other minds or an external world, must admit there are words a person can learn only from his own sensory experience (Goldstein 1996: 141).

Much equivocation is to be found in this passage. If, for example, 'learning through your own sensory experience' means what Wittgenstein's 'private linguist' means, namely, being confronted by an 'experiential Given' which one goes on to name 'privately', then this is not an 'innocent truth' at all, but rather a substantial metaphysical claim. That Goldstein thinks the thesis innocuous perhaps stems from the fact that he is conflating two different things: 'to perceive something' and 'to have a sensory experience'. Although the two phrases sound similar, they are evidently not the same, something clearly revealed by the fact that Goldstein is committed to the further claim that one perceives *only* one's sensory experiences (and not, say, physical objects):

Yes, we learn ‘red’ in a way empirically the same as the way you describe. We learn ‘red’ through the kind of experience you describe as one in which we encounter ‘an external object which our parent calls ‘red’. However, when a person has this type of experience, he is directly aware of only his own sensations. A person gains knowledge of an external world only by reasoning from these sensations to their external cause (Goldstein 1996: 145-6).

The confidence with which Goldstein asserts that one is only ever ‘directly aware of one’s sensations’ and, hence, that knowledge of ‘external objects’ can, at best, be inferential, betrays Goldstein’s antecedent (and perhaps unwitting) commitment to the Myth, since if he didn’t reify his ‘sensations’ into ‘intermediary’, ‘self-subsistent’ items intervening between observers and the external world, what would entitle him to make this claim? How does he know, for example, that we do not perceive physical objects given that this is how things actually seem to us? It is surely neither an *a priori* nor an empirical truth that we perceive *only* our own sensations²⁷. Consequently, if one doesn’t already *start* with the idea that there must be an interface (‘epistemic intermediaries’), what reason would one have to doubt that one is mostly in touch with a physical reality?

Perhaps Goldstein is taken in by what, in the first section, I have called ‘aggregate arguments’ – i.e. to think that it follows from the fact that one can sometimes be misled about what one perceives, that one could always be misled about what one perceives, and, hence, that one may never be in touch with ‘external reality’ at all, but only with a pale ‘copy’ of it –

²⁷ In other words, to have sensations may be causally necessary to perceive physical objects, but this does not imply that one perceives *only* one’s sensations and not physical objects.

‘sensory Experience’ with a capital ‘E’, the object of the Myth²⁸. But we have already seen that such arguments are fallacious, and not by themselves sufficient to generate either the cogency of NET, or the radical sceptical scenario. Consequently, Goldstein’s objection fails.

Finally, as an option of last resort, a sceptic may want to bite the bullet and say something along the following lines: Let’s grant that you have successfully established that the BFB cannot succeed in giving meaning to the expressions of her private language, because she cannot establish on her own conditions of correct use for these expressions. How does it follow from this that our perceptual experiences cannot be mere appearances? All that follows is that, if that’s what they are, then there are no conditions of correct use for the terms with which we describe our experiences. But none of this shows that this isn’t the situation we find ourselves in. Remember that the BFB may well be under the illusion of having established conditions of correct use. Why can’t that be our situation?

The answer is: If the BFB is under the illusion of having established conditions of correct use, then ‘her’ ‘words’ mean nothing for they have no conditions of correct use. And if ‘her’ words mean nothing, then the BFB is not ‘saying’ anything. Consequently, the radical sceptic cannot even ‘save the appearances’. That is to say, the sceptic’s ‘appearances’ – never mind whether they actually are of an ‘external reality’ or not – do not possess *any content* (that is, they’re not even *appearances*) and, hence, everything goes dark in the sceptic’s ‘interior’. Since the radical sceptical scenario, however, is the contention that all of the appearances as of an external world may be nothing more than *mere* appearances (i.e. just episodes of one’s consciousness), if it turns out that the sceptic even requires us to deny that such ‘mere appearances’ are possible, then there can be no such thing as radical scepticism either, and the sceptic has sawn off the branch on which she was sitting.

²⁸ In a recent book, Searle calls this the ‘Bad Argument’ against direct realism: ‘One of the biggest mistakes in philosophy in the past several centuries’ (Searle (2015: 10)).

IV Conclusion

If what I have argued in this paper is right, then the radical sceptic (or our ‘sceptical’ *alter ego*) is faced with an insuperable problem: if we accept the view that knowledge of the facts may never be possible, then, *pace* McGinn et al., we cannot immunize knowledge of the content of our sense-experiences against radical sceptical doubt either. This leaves it impossible to explain – without appealing to incoherent notions such as the Myth and Magical Theories of Reference – how our putative ‘external world’ beliefs can have the conceptual content that they do at all. Given that we *are* able to formulate such propositions, however, knowledge of the facts, must, in principle, be possible²⁹.

Genia Schönbaumsfeld, University of Southampton

gmes@soton.ac.uk

References

Ayer, A.J. (1971), ‘Could Language be Invented by a Robinson Crusoe?’ in O.R. Jones (ed.), *The Private Language Argument* (London: Macmillan), 50-61.

Boghossian, P. (2002), ‘The Rule-Following Considerations’ in Alexander Miller and Crispin Wright (eds), *Rule-Following and Meaning* (Chesham: Acumen), 141-87.

²⁹ I would like to thank Guy Longworth, and the audiences at the Aristotelian Society and the UEA Wittgenstein workshop for helpful comments and suggestions on an earlier version of this paper.

Davidson, D. (1986), 'A Coherence Theory of Truth and Knowledge' in Ernest Lepore (ed.), *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson* (Cambridge: Blackwell), 307-19.

Davidson, D. (2001), 'On the Very Idea of a Conceptual Scheme' in *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press), 183-99.

Descartes, R. (1641/1986), *Meditations on First Philosophy*, trans. John Cottingham (Cambridge: Cambridge University Press).

Fogelin, R. (1987), *Wittgenstein* (London: Routledge and Kegan Paul).

Goldstein, I. (1996), 'Ontology, Epistemology, and Private Ostensive Definition', *Philosophy and Phenomenological Research* 56 (1):137-47.

Hervey, H. (1971), 'Private Language and Private Sensations' in O.R. Jones (ed), *op.cit.*, 76-95.

Jarvis Thomson, J. (1971), 'The Verification Principle and the Private Language Argument' in O.R. Jones (ed.), *op.cit.*, 183-204.

Kenny, A. (2006), *Wittgenstein* (Oxford: Blackwell).

Kripke, S. (1982), *Wittgenstein on Rules and Private Language* (Cambridge MA: Harvard University Press).

Malcolm, N. (1971), 'Exposition and Criticism of Wittgenstein's *Investigations*', in O.R. Jones (ed.), *op.cit.*, 33-49.

McDowell, J. (1994), *Mind and World* (Cambridge MA: Harvard University Press).

- McDowell, J. (2003), 'Subjective, Intersubjective, Objective', in *Philosophy and Phenomenological Research* 67 (3): 675–81.
- McGinn, C. (1986), 'Radical Interpretation and Epistemology' in Ernest Lepore (ed.), *op.cit.*, 357-70.
- Mundle, C.W.K (1971), 'Behaviourism and the Private Language Argument' in O.R. Jones (ed.), *op.cit.*, 103-17.
- Putnam, H. (1981), *Reason, Truth and History* (Cambridge: Cambridge University Press).
- Putnam, H. (2002), *The Collapse of the Fact/Value Dichotomy and Other Essays* (Cambridge MA: Harvard University Press).
- Schönbaumsfeld, G. (2016), *The Illusion of Doubt* (Oxford: Oxford University Press).
- Searle, J. R. (2015), *Seeing Things As They Are* (New York: Oxford University Press).
- Sellars, W. (1956/97), *Empiricism and the Philosophy of Mind* (Cambridge MA: Harvard University Press).
- Strawson, P.F. (1971), 'Exposition and Criticism of Wittgenstein's *Investigations*' in O.R. Jones (ed.), *op.cit.*, 27-33.
- White, R. (2014), 'What is my Evidence that here is a Hand?' in D. Dodd and E. Zardini (eds.), *Scepticism and Perceptual Justification* (Oxford: Oxford University Press), 298-321.
- Wiggins, D. (1980), *Sameness and Substance* (Oxford: Blackwell).
- Wittgenstein, L. (1965), *The Blue and Brown Books* (New York: Harper and Row).
- Wittgenstein, L. (1969), *On Certainty*, ed. G. E. M. Anscombe and G. H. von Wright, trans. Anscombe and D. Paul (Oxford: Blackwell).

Wittgenstein, L. (2009), *Philosophical Investigations*, ed. P. M. S. Hacker and J. Schulte, trans. Anscombe, Hacker, Schulte (Oxford: Blackwell).