



Naturalised Inferentialism and the Incompleteness Problem

Jaakko Reinikainen¹

Accepted: 20 May 2024
© The Author(s) 2024

Abstract

The paper argues that the naturalised version of semantic inferentialism advanced by Jaroslav Peregrin faces a problem which, following Michael Devitt, I call the incompleteness problem. The main issue has to do with how, according to inferentialism, language is connected to the world. My main claim is that Peregrin's Protagorean account of correctness is in tension with the idea, made also by Robert Brandom, that language is embodied in the world analogically to how physical objects are embodied in games like football. Against this, I show the two are in fact importantly disanalogical. To solve the incompleteness problem, I argue that naturalised inferentialism should learn the central lessons of semantic externalism, namely that the connection between language and the world must be fundamentally external to the mind, or in Peregrin's case, to the society of normative attitudes.

Keywords Inferentialism · Meaning · Semantic externalism · The incompleteness problem · Naturalism

1 Introduction

This paper has two aims. First is to argue that the naturalised version of inferentialism advocated by Peregrin (2014a, 2024) faces a problem concerning the determination of linguistic meaning. The second aim is to show that the problem essentially emerges from troubles in understanding, in inferentialist terms, how the world is 'embodied' in language. To solve the problem a viable recourse could be, I argue, to an independent theory of reference, particularly the causal-historical theory of reference.

The argument owes inspiration to the problem of incompleteness which Michael Devitt has raised against descriptivist theories of meaning and reference (Devitt 1996; p. 159).¹ While I am not claiming that Peregrin's inferentialism is an instance of descriptivism, I do believe the two isms share certain traits, including the problem of incompleteness.

I present my case as follows. First, I offer a very focused exposition of Peregrin's Protagorean account of normativity

along with the embodiment claim. Next, I show how this account faces the incompleteness problem. Finally, I argue that a ready way to solve the problem is by admitting some semantic externalist elements into the account.

1.1 Peregrin's Inferentialism: An Overview

The main idea of semantic inferentialism is that meanings can be modelled by, and to some extent identified with, inferential roles. Peregrin's (2014a; 2024) brand of inferentialism takes inspiration from Robert Brandom's groundbreaking work (1994), but with certain important differences. In this subsection, I briefly elaborate on the main idea of inferentialism as advanced by Brandom, then show how Peregrin's version deviates from him.

Inferentialism in its most general sense is a species of so-called use-theories of meaning, which broadly speaking seek to understand and explain the phenomenon of meaningful language in terms of speakers' linguistic behaviour. However, for a proper understanding of the phenomenon, it is not sufficient to examine actual, observable patterns of use, nor even abstract dispositional patterns of use, but rather patterns of (semantically) correct use. The idea is broadly based on later Wittgenstein's (1958) remarks about how language can be viewed as analogical to many games, which crucially involve the distinction between the rules and the moves. So, very bluntly, meanings are similar to

¹ For descriptivism, see Kripke (1980) and a comprehensive discussion by Raatikainen (2020).

✉ Jaakko Reinikainen
jaakko.reinikainen@tuni.fi

¹ Tampere University, Faculty of Social Sciences, Philosophy, Tampere, Finland

rules of use. The way inferentialism distinguishes itself as a use-theory of meaning is with the claim that the paradigmatic kind of use consists of drawing certain types of inferences, called by Brandom (after Wilfrid Sellars) ‘material inferences’ (1994, 97).

A key theoretical motive for inferentialism comes from discontent with representationalism, i.e. the perceived failure to give substantial content to the intuitive idea that language is meaningful in virtue of representing the world (Brandom 2001, 7–10). While Brandom, and to some extent Peregrin, allow that language does have a representational dimension of some kind, the theoretical point of inferentialism is to offer an alternative, non-representational explanation of the phenomenon of meaningfulness.

Neither a single paragraph nor a hundred will do true justice to Brandom’s intricate, fine-tuned development of inferentialism. The point I want to focus on here concerns the status of ‘correctness’ that plays a fundamental part in the theory, and which from the start gathered some of the most pointed criticism (Rosen 1997; McDowell 1997). A key problem about correctness at focus in this paper is how it emerges in nature. The analogy to games stresses rather than illuminates the problem, for the normativity of their rules is clearly dependent on the powers of our intentionality and willingness to follow and play by the rules that we have devised.

To be more precise, there are two important disanalogies between ordinary games and the language game which any inferentialist account must address. First is that the rules of games are usually explicit, explicitly (albeit perhaps only partially) known by the players, and often explicitly cited by them when playing the game. This is not the case with the language game, where the rules are *primarily* implicit, known implicitly, and citing them (or rather, making them explicit) is not too straightforward as it is with games. The second point is that the normativity of the rules of games is (usually) wholly dependent on intentions of the players which are, again, expressed explicitly (For example by saying, ‘Let’s play chess.’) Not so with the language game. For example, Brandom’s ‘Queen’s Shilling model gives the primary power of binding the speaker to the norms to others around her (1994, 162).

A third disanalogy is that games are as a rule invented by someone, yet no one invented language in that sense. So, if meanings are analogical to rules, they must have grown from ‘bottom up’, not handed top-down. But how could such a remarkable thing have happened, even in theory?

A key complaint of Peregrin against Brandom is that, at least in *Making It Explicit*, he does not really give an account of this developmental problem. Moreover, the reason for the lack is that Brandom has strong *sui generis* understanding of semantic, or more broadly discursive, normativity, viewing

it as something irreducible to non-normative grounds (Brandom 1994; p. 627). This point, in turn, threatens to make normativity something, if not super-natural, then at least not fully explicable by the natural sciences. Peregrin has no such reservations about normativity: it is at bottom a fully natural phenomenon, and the key to it are the normative attitudes of speakers (Peregrin 2024, 3).

It is interesting to note that, as a supporter of naturalism in the philosophy of language, Peregrin shares much in common with a representationalist like Devitt (1996). The common thread of their (programmatic) commitments to naturalism is that philosophical theories about language and meaning should not only seek to be compatible with the special sciences, but actively engage with their methodologies and discoveries. Their differences stem not so much from what naturalism means but from what needs to be ‘naturalised’. For Peregrin, what primarily needs naturalising are the moves in the language game of giving and asking for reasons, as well as the ontogenetic and phylogenetic history of language acquisition (see especially his 2014b). For Devitt, what primarily needs naturalising is the representational relation of language to the world (1996, 2–3).

It is Peregrin’s naturalised notion of semantic correctness which this paper aims to problematize, but not so much from the angle of naturalisation, as on the grounds of more traditional semantic externalist arguments. Briefly, the main idea is that while I view the aim of naturalising meanings as a good one, in order to do that, inferentialism alone will not be enough insofar as it drops reference from its theoretical toolbox. To explain why that is, we must next examine in detail Peregrin’s naturalised account of semantic correctness.

1.2 Protagorean Normativity

Already in his early work on inferentialism, Peregrin (2000, 8) recognized, following Devitt and Sterelny (1987/1999), that losing reference from the explanatory toolkit risks viewing language too much as a game of chess, i.e. a self-contained system that is only thinly constrained by how the world is, if at all. The generic task for retaining the world, or ‘semantic objectivity’, is then to explain how the game of giving and asking for reasons can be suitably constrained by the way the world is. This task Peregrin hands over to the notion of ‘Protagorean normativity’:

“I argue that the most primitive form of correctness is Protagorean in this sense: in these most primitive cases, something is correct iff it is held for correct by a vast majority. [...] And, moreover, I am convinced that any normativity ultimately derives from this primitive, Protagorean form. (Peregrin 2024; p. 105)”.

Peregrin then introduces an important specification to the idea of Protagorean normativity: any actual set of normative attitudes may in some later time turn out to be incorrect. So, since any actually held set of normative attitudes can be later found incorrect, no such set can be the Protagorean bedrock of normativity. Rather, the idea proper is that:

what is decisive would be the normative attitudes that would result from an exhausting research and faultless argumentation, which is not something to be encountered in the real world. It is, then, obvious that any *de facto* posterior attitude may be mistaken – there is always the possibility that it will still be altered. (Peregrin 2024; p. 106)

In other words, the bedrock of normativity proper is an epistemically idealised set of normative (majority) attitudes that agree on what is correct. As Peregrin points out, however, even idealised agreement in attitudes is not yet objective correctness in the sense of truth, which he agrees can in principle diverge from agreement in attitudes formed even in idealised epistemic circumstances (Peregrin 2024; p. 104). The way objectivity is meant to arise from Protagorean normativity happens by distinguishing between the criteria set by the attitudes and facts about whether the criteria are met in actual cases. In schematic form:

if a property *P* is Protagorean, then *to have P* is simply *to be taken to have P*. If it is not, then *to have P* is something else, let us say *to display a feature F*. In a particular case, people may be mistaken about something displaying *F* and it may take some research to find this out; it may be that after the research is carried out, the opinion of the community changes so that the thing in question then *has P* just in case it *is taken to have P*. Thus, what is decisive are not the prior attitudes (before the research), but the posterior ones (after the research). But in order for the research to effect this, it is necessary that there is an agreement that *to have P* is *to display F* and that hence displaying *F* is the criterion of having *P*. In the simplest case, this is Protagorean. (Peregrin 2024; p. 107)

As an example, take the majority judgement that Lassie is a dog, i.e. has *P*, because she displays feature(s) *F*. The majority can be mistaken whether Lassie really displays features *F*. If they are mistaken, it is because in some epistemically idealised posterior set of majority attitudes, it is found out that Lassie does not display *F*, which is an agreed upon necessary criterion for having *P*, or being a dog. This latter point is not what the idealised majority can be mistaken about, according to Peregrin, since it is their (majority) agreement

which immediately settles the matter. Furthermore, the idealised majority attitudes also immediately determine the correct methods for deciding whether *F* is displayed in any actual case:

Hence, in the end, the fundamental agreement of normative attitudes concerns not what is correct, but rather how to find out what is correct – not the property itself, but rather its criterion. We agree that something is a dog iff it displays such and such features. (Peregrin 2024; p. 112)

This summarises the initial view of Protagorean correctness. One point that I want to emphasise early on about the account is that it is committed to what I shall call *the strict division of labour* between the attitudes and the world. Simply put, the division of labour is such that the attitudes determine, in the ideal limit, the correct applicability criteria for any predicate, while it is up to the (non-attitudinal) world which criteria are met in any actual case.

Next, I focus on another, related idea in Peregrin that is supposed to help solve the problem of how the world can appropriately constrain the game of giving and asking for reasons: embodiment.

1.3 Embodiment

The previous section sought to show Peregrin's commitment to what I called the strict division of labour between the attitudes and the world. Now, it's important to focus on another key idea of Peregrin's which is meant to balance out the strict division of labour. For what the strict division of labour seems to immediately entail is that attitudes alone can determine meanings in the same way that they determine the rules of chess. This may leave it mysterious how our language can include genuinely empirical content, and not be left 'spinning in the void', to use John McDowell's famous phrase.

This balancing commitment I call the *embodiment claim*. The main idea, I believe, comes from Brandom, who also calls it 'incorporation':

Solid discursive practices incorporate nonlinguistic things in them (are *corporeal*). In the same way, the practice in which a performance can have the significance of hitting a home run incorporates objects. (Brandom 1994; p. 632, fn.)

Peregrin, too, frequently appeals to the difference between abstract and physical games like chess and football to substantiate the embodiment claim (e.g. 2014a, 34, 36). One way to understand the difference between abstract and

physical games is that chess does not necessarily involve any physical equipment, whereas football does.² How is speaking a language like playing football then? What is the necessary physical equipment for speaking, say, English? Peregrin answers:

Now it is important to realize that in the case of our language games, the whole world is our (potential) shared equipment. Just as when playing chess we are not moving pieces of wood, of which we know that they play the roles of rooks and bishops, but are playing directly with the rooks and bishops, so when playing our language games we are living among the things constituted by these games. We structure the world by means of concepts cast in the mold of the game of giving and asking for reasons into the shapes of parts of propositions. This is not to say that these things are ‘imaginary’ or ‘unreal’ – no more than rooks and bishops are unreal. It is to say, though, that they are what they are only in the context of our language games. And insofar as we are ‘inside’ the games, this is what they are *ultimately*. (Peregrin 2014a; p. 112)

This citation implicates the division of labour introduced above. It is the job of the rules, hence attitudes, to conceptually constitute bits of the physical world as ‘dogs’, ‘cats’, etc. in the same way it is the job of the attitudes to conceptually constitute bits of wood and leather as ‘rooks’, ‘kings’, ‘footballs’, etc. It bears emphasis that Peregrin seems to take the analogy to games very literally here: it is more less in the same sense that attitudes decree the correct shape, size, and weight of a football that they also decree the correct Protogorean properties for something to count as a ‘dog’, a ‘cat’, etc. The world by itself does not have any *conceptual* credentials or authority; the individuation of bare physical stuff into distinct kinds only happens with the advent of normative attitudes.

To take stock, I have now attributed two key commitments to Peregrin regarding the fundamental nature of semantic correctness and the relationship between language and the world:

1) There is a *strict division of labour* between the attitudes and the world such that the attitudes determine (in the ideal limit of inquiry) the applicability criteria for predicates whereas the world determines whether they are fulfilled in any particular case.

² We can finetune this point with two quick remarks. First, one could play chess only ‘in the head’, which is not possible with football. If the mind isn’t physical, then perhaps one could play chess even without needing a body. Second, even where physical equipment is used to play chess, this equipment is arbitrary in constitution, limited only by practical considerations. Not so with football: to be kicked about at all, a football must have, e.g., certain elasticity and weight.

2) However, the attitudes and the world are not ontologically speaking fundamentally different structures, but rather the attitudes and language have grown to fit the world, *embodying* parts of it (Peregrin 2024; p. 206).

In the next section, I argue that there is unresolved tension between these two commitments as shown by the incompleteness problem. In order to make good for the embodiment claim, the strict division of labour must be negotiated. This, in turn, will have important consequences for the Protogorean account of correctness.

1.4 The Incompleteness Problem

To start off, there is an obvious way in which the analogy between football and language breaks down. To investigate the correctly attributed properties of a football, it would be ill-advised to investigate actual, concrete objects that are treated as footballs. The reason is that any one of them, or all of them, could be faulty in some respects: it’s not the footballs that make the rules for football, after all. To investigate the correct properties of a football we must consult the rules of football, and thus ultimately the normative attitudes of the people who play football. But this is clearly *not* analogous to how we investigate the correct properties of cats, i.e. the properties that should be ascribed for an object to be properly called a cat. Instead, we must investigate actual cats and what properties they have. So, cats are not equipment for our language games in the way footballs are equipment for football in that the cats make their own rules about what properties are correctly ascribed to them!³

In other words, I think we can safely dispense with the *literal* analogy that Peregrin, following Brandom, draws between (empirical) language and physical games. But what, then, is the analogy supposed to show? Since it is not literally true that attitudes decree the properties that are correctly ascribed for cats, in what sense are cats part of the ‘equipment’ for English? In what sense do the attitudes conceptually constitute chunks of brute physical matter as ‘cats’? This is where I think the embodiment claim remains underdeveloped.⁴ I return to this point later once we have gone through the problem of incompleteness.

Let’s continue by revisiting the Protogorean account of correctness. First of all, I take it that a set of attitudes is epistemically idealised in Peregrin’s sense when the attitudes would no longer change no matter how much argumentation

³ Naturally, Peregrin would protest that the correctly attributed cat-properties also depend on rules, only in the feline case these are instituted implicitly. The point I wanted to make here is that the analogy between games and empirical language, if interpreted literally, is far more radical than I think Peregrin lets on.

⁴ Elsewhere, I have argued that Brandom implicitly develops the embodiment claim via the idea of conceptual realism, i.e. that the world itself is conceptually structured (Reinikainen 2021).

or new empirical information is added. This is why no actual set of attitudes is ever Protagorean in the ultimate sense: because actual attitudes are always potentially open to change in the light of new arguments or information. For one thing, this leaves Peregrin's proposal open to foundational epistemic scepticism leading to foundational semantic scepticism. If we don't know which actual attitudes of ours would in their epistemic limit stabilise, i.e. experience no further change, or reach majority agreement, then we don't know whether *any* one of them would ever stop changing or reach majority agreement, i.e. have an ideal limit. But if there is no epistemic limit where adding new arguments or information would not lead to further change in attitudes, there are neither stable, unchanging semantic criteria determined by the attitudes. But in that case, there are no facts about, e.g., whether Lassie is correctly called a dog or not because there are no stable criteria (or methods for finding the criteria) that would settle the matter.

The mere possibility of foundational scepticism is not very threatening for Peregrin, of course. In any case, such scepticism does not pose a problem unique to him. But noting the possibility of foundational scepticism leads us to see the more serious problem in Peregrin's Protagorean account. The difficulty is that for Peregrin, the correct methods for determining the criteria for being P must come *before* it is determined which objects are (candidates for) being P. However, this is arguably not usually the case, especially not with empirical terms like 'dog'.

We can present the problem by returning to the foundational sceptical problem from another angle: why is it that the idealised majority attitudes would settle for one set of methods (and/or criteria) as opposed to another for being P? For example, why would they hold that barking is a criterion for being a dog, or that finding out that Lassie barks is one method of finding out that she is a dog?

I think Peregrin has two possible answers available here. My aim is to show that one of these conflicts with his inferentialism, while the other leads to some arguably implausible conclusions.

To clarify, the assumption now is that the majority attitudes in the ideal limit of inquiry would ultimately reach a point where no further change occurred (i.e. 'converge'), rather than continuing forever in response to new arguments or information (i.e. 'diverge'), which would lead to indeterminacy. So, we assume that foundational scepticism is false. The point we want explained is why, in principle, the majority attitudes would settle for one set of criteria for being P rather than another, assuming they do stop at some definite set? The simplified example I continue to use is why is barking a Protagorean criterion for being a dog.

A natural response might be: because (only) dogs bark. But this cannot be Peregrin's answer because it would lead

to a quick circularity. The answer to why barking is a Protagorean criterion for being a dog cannot be that (only) dogs bark because this is exactly what we want to know, i.e. why is there a Protagorean connection between dogs and barking? But the first answer simply restates that barking is a property (only) of dogs. An immediate remedy is close at hand, though: barking is a Protagorean criterion for being a dog because (only) *these kinds of objects* bark, where 'these kinds of objects' implicates an ostensive definition of pointing at some dogs.

But we can further ask, why should the fact that (only) *these kinds of objects* bark have anything to do with the criteria for being a dog? And Peregrin's answer cannot be 'Because (only) *these kinds of objects* are dogs', for that would lead to the earlier tautology via a little detour. To see this, consider the following dialogue:

1. Why is barking a Protagorean criterion for being a dog?
2. Because (only) *these kinds of objects* bark.
3. Why is their barking relevant for question (1.)?
4. Because (only) *these kinds of objects* are dogs.⁵

If we swap the response on line number 4 to line number 2, we arrive at the original circularity:

1. Why is barking a Protagorean criterion for being a dog?
2. Because (only) dogs (i.e. *these kinds of objects*) bark.

What exactly goes wrong in these responses? The fundamental problem, already mentioned above, is that in Peregrin's account, the Protagorean criteria for being P must come before any actual objects are recognized or treated as being P. The job description of Protagorean norms is to establish an ultimate set of criteria for the applicability of any predicate. So, the explanation for why the majority attitudes in the ideal limit would settle for one set of criteria over another cannot in turn rest on further, already established criteria. They cannot, for example, assume the truth of the claim that only dogs bark because such a truth is precisely what their function is to establish, by establishing the Protagorean criteria or methods for the applicability of the predicate 'is a dog'. But how can we resolve the question of why the attitudes would settle for one set of criteria as opposed to another in the case of an empirical concept like 'being a dog' without *using* the very word the applicability criteria for which we are seeking, which constitutes the first, natural answer available to Peregrin? It bears emphasis that this is not an epistemological problem: the problem is not that we don't know how to describe the correct epistemic idealisation. Rather, the problem is to show what fact

⁵ It would not change much to change line 4 to 'Because (only) *these kinds of objects* are treated (or recognized) as dogs.' We would still be using the word 'dog' in explaining criteria for being a dog, which would be circular.

or reason *in principle* would make the attitudes in the ideal limit settle for one set of criteria as opposed to another.

Here is where the second possible response by Peregrin comes in. Simply put, there is no reason why the majority attitudes in the ideal limit, supposing they would stop at some definite set of Protagorean criteria for being a dog, would settle for barking. They simply do stop somewhere, and where exactly they stop is arbitrary, or perhaps conditioned in complicated ways by human nature, if there is such a thing.⁶ In any case, it is a mistake, or at best an irrelevance, according to this response, to ask why the attitudes would settle for one set of criteria as opposed to another. They just do.⁷

Next, I will explain why the first, natural response ends up contravening Peregrin's inferentialism, while the second leads to implausible conclusions.

To repeat, in Peregrin's account, the applicability conditions of predicates are set by Protagorean criteria established in the ideal limit of inquiry *before* any actual objects are identified as falling under the predicate. In other words, the criteria come before the reference of words, because the criteria are meant to (partially) determine reference, or rather, correct applicability. This is to be expected, of course, because the whole point of inferentialism (as Brandom and Peregrin pursue it, anyway) is to oppose giving reference an independent, autonomous explanatory role (Peregrin 2014a; p. 24).

However, the incompleteness problem pushes us to see why getting rid of reference is difficult. We can see this by showing an easy solution to the problem. Simply put, we give up the idea that the Protagorean criteria determine the applicability conditions of the predicate 'is a dog'. True, barking is a criterion we use to identify dogs, but it is not what determines whether an object *is* correctly called a dog, or what the reference of 'dog' is. Rather, the reference of 'dog' is established independently of the criteria, which thus serve primarily an epistemic, not a semantic, role. So, the short reason why the first response ends up contravening an important aspect of Peregrin's inferentialism is that it naturally gives reference an independent, important explanatory role. Next, let's turn to the second response and how it might avoid this outcome.

⁶ To get an idea of what this might mean, consider the physical limits which human biology sets on the rules of football: the games cannot last several days, the field can't be kilometres long, the ball can't weigh 100 kilos, etc. Subtler reasons would appeal to facts about why we find the activity of kicking a ball fun.

⁷ A third possible response would be to say that the Protagorean criteria for being a dog, whatever they may be, are analytically true. But I don't think Peregrin, as an admirer of Quine (Peregrin 2024; viii), would find this response tempting. In any case, it's hard to see how the applicability criteria of 'is a dog' could possibly be analytic.

The second response basically amounts to a form of primitivism by denying that we could, or would need to, have an explanation for why the majority attitudes in the ideal limit would settle on one set of criteria as opposed to another for a given predicate. This is not to say that the matter is completely arbitrary: perhaps human nature or some other generic background condition plays a role here. Still, in spirit, this response leans more towards rejecting the question than meeting it head on. A ready way to evaluate this response, then, is to evaluate how reasonable the question is.

On my part, I think it is extremely reasonable. The second response essentially doubles down on the idea that all 'correctness', all 'normativity', originates from the attitudes of rational, speaking subjects. But isn't it reasonable to think that sometimes, also the world can function as a 'standard of correctness' independently of the attitudes – simply by being the way it is? True, Peregrin does give credit to the world as a standard of correctness, but only by understanding the attitudes (and thus rules) as *embodied* in the world, *incorporating* bits of it (2014a, 15, 113). In this way, the world itself limits the arbitrariness of the attitudes, though not because language represents (or 'mirrors') the world, but because it 'is bound to "fit" into it' (Peregrin 2024; p. 206).

However, in the previous section I showed that Peregrin's claims about incorporation and embodiedness are not meant to contest the claim that the world does not set standards of correctness by itself. The world may, in complicated ways, *influence* the Protagorean majority attitudes in the ideal limit (in the same way human biology influences the rules of football), but strictly speaking the world may not decide for the attitudes. Humans could rule that football games last a week, however counterproductive that would be. The attitudes are supposed to be sovereign in this sense. Moreover, at the beginning of this section, I pointed out that it is difficult to see how Peregrin's main way to substantiate the embodiment claim – the equipment analogy – could be literally true of a predicate like 'is a dog'. Next, I further press the argument above to support the disanalogy, and thus the conclusion of why not all fundamental normativity can be *only* up to the Protagorean attitudes even in the second reading that approaches primitivism. All this is ultimately meant to show the inherent tension between Peregrin's commitments to the strict division of labour and the embodiment claim.

To continue the argument, notice first that the Protagorean criteria that determine what it is to be P are descriptive. While the criterion of barking for being a dog is an obvious example, the point can also be seen in that the ultimate criteria are supposed to be reached by faultless argumentation

in the ideal limit of inquiry. How else could that be achieved unless the criteria were descriptive?

Equally obvious is the point that the descriptive criteria must be meaningful. But how is the meaning of the criteria for being P determined? For example, how is it settled what are the criteria for an object ‘barking’? Perhaps this, too, can be determined by descriptive criteria reached by argumentation. But then the same point that arose for being a dog meets the criteria for barking: why does making *that kind of sound* have anything to do with ‘barking’?

The point of the incompleteness problem is that this last question cannot be settled by further criteria *ad infinitum* without lapsing into the foundational epistemic scepticism. Equally unlikely is that the bottom criteria are analytically true.⁸ At some point, the answer seems to be that we simply refer to *that kind of thing* with this type of word. But if that is true, the reference (i.e. correct applicability) of some words must be secured independently of any debatable, descriptive criteria.

But perhaps the incompleteness problem misses something important, namely that the bottom criteria are not, in fact, descriptive, but rather ostensive? While the criterion of barking is descriptive, the criteria for a sound to count as barking need not themselves be descriptive, but rather ostensive and rely only on perceptual recognition by majority attitudes in the ideal limit.

This proposal seems overwhelmingly plausible, and it is, of course, not a new one. Wittgenstein famously claimed that the possibility of a shared language presupposes agreement not only in definitions, but also in judgements (Wittgenstein 1958, § 242). What he probably meant was not agreement in all judgements, but only in some set of perceptual ones like something being green or round or sounding like barking: Crispin Wright calls these ‘basic judgements’ since they involve ‘basic concepts’ (Wright 2001, 59). The point that I want to raise here is that if the bottom criteria are ostensive rather than descriptive, that seems to imply an important switch in Peregrin’s Protagorean account of correctness. We recall from above that in the initial presentation of the account, there is a strict division of labour such that the criteria are set by the attitudes only while it’s only up to the world whether they are ever fulfilled. But in the case where the bottom criteria are ostensive, the division becomes muddled: a basic concept like barking is identified by the majority attitudes by its *paradigm instances*, which must be *actually* perceived. So here, the general criteria do

not come before actual, particular instances of application. Rather, it seems that the attitudes and the world together set the standard for basic concepts: it is *these* attitudes in conjunction with *these* paradigmatic instances that set the criteria for something to sound like ‘barking’ in some actual, non-ideal context.

Perhaps this conclusion is agreeable to Peregrin; perhaps it is one way how the embodiment claim can be cashed out. But in that case, reference plays a more important role in the Protagorean account than he lets on. The criteria for basic concepts are set by the majority attitudes, not in the ideal limit, but in some *actual* context where they are able to recognize a certain kind of sound as ‘barking’, some actually seen wavelength of light as ‘green’, etc. Peregrin’s emphasis on specifically posterior attitudes for determining the correct criteria (and methods for finding them) becomes redundant in this picture, since the only way to identify the criteria is by ostensive reference to some actual, paradigmatic examples that are as readily available to present attitudes as they are to later ones. Yet without admitting that the criteria for basic concepts are non-descriptive and ostensive, it is difficult to see how he can avoid the incompleteness problem, and by that token the implausible conclusion of foundational scepticism.

How is it that the ideal limit of inquiry is redundant for determining specifically ostensive, non-descriptive criteria of application? After all, there is a clear sense in which perception, for example auditory perception, can be ideal or non-ideal. It is idealised when we abstract away all confounding factors, such as extraneous noise or an internal organic failure, and it is non-idealised when such confounding factors may play a part in the ostension. Notice, however, that the idealisation of perception requires descriptive criteria to determine what counts as a confounding factor, what authentic, well-functioning perception. So, where do *these* criteria come from; how do *they* become meaningful? The answer that appears unavoidable is that the bottom correctness must rest with some non-ideal, actually instantiated base of applications. It is the sound as heard in nature, as opposed in the ideal limit, that gives the paradigmatic standard for barking.

There is an objection which the inferentialist could make in response to the incompleteness problem that seems quite natural, but which I think is based on confusing the incompleteness problem with another issue. The key to the objection is the idea that the meaning-conferring norms are supposed to be *implicit* in practices. So, the idea goes, the criteria that determine correct applicability need not be explicitly thought of by the participants in the ideal limit of inquiry – this would commit the fallacy peculiar to what Brandom has called ‘regulism’, which identifies all rules with explicit precepts (Brandom 1994, 20).

⁸ Here, I am not presenting an argument for the bottom criteria not being analytic. My reason for dismissing this option is that Peregrin is not tempted by it: ‘The last remnants of my belief that there are some unquestionable analytic truths were shaken when I saw, in the window of a gift shop, small wooden ovals advertised as “wooden stones”’ (Peregrin 2014, 27).

But the incompleteness problem is not about the (implicit or explicit) *status* of the norms and criteria that determine correct applications. Rather, the problem is that the criteria must be *descriptive* in order to determine correctness in the ideal limit. But if they are descriptive, the terms in the description must eventually have their meanings determined by some other means than by more descriptions – otherwise inferentialism remains incomplete. The main lesson of the incompleteness problem is that there must be a non-descriptive way to determine the ‘correctness criteria’ of some applications of words. A ready answer to what that means is that the bottom criteria are ostensive. But in that case, as we saw, the ideal limit of inquiry becomes redundant, and the ‘sovereignty’ of the attitudes comes to be shared with the world, breaking pure Protagorean correctness.

1.5 Lessons of the Incompleteness Problem

The discussion above can be summarised in three distinct moments:

(1) Peregrin’s Protagorean account of correctness rests on a strict division of labour between the normative attitudes and the world. On the one hand, the majority normative attitudes set the criteria (as well as the correct methods for establishing the criteria) for the applicability of predicates in the ideal limit of inquiry. On the other hand, the world determines which of the criteria are fulfilled in any actual instance.

(2) The incompleteness problem brings the strict division of labour under pressure. The applicability criteria of basic concepts are not set in the ideal limit, and not solely by the attitudes, but rather by the attitudes in conjunction with the world in some actual, non-ideal context.

(3) Consequently, the ideal limit of inquiry becomes redundant for determining the correct application of basic concepts.

In this section, I draw certain general conclusions from the incompleteness problem. The main moral is the semantic externalist idea that it is very difficult, and arguably impossible, to explain how language can be *about* the world (i.e. suitably constrained by it) without admitting some attitude- or mind-external relation to bridge the gap. In programmatic terms, this is exactly what Peregrin acknowledges: language is a natural phenomenon that has essentially grown to fit the world or be embodied in it. The problem is that what appears to be his primary way to substantiate this claim – the equipment analogy combined with the Protagorean account of correctness – is insufficient to explain the aboutness relation. There is a *logical* gap between the Protagorean criteria set by attitudes and the particular instances of application which the criteria are meant to settle which allows us to ask, why is it that *these* criteria are *about* this property or

predicate? This is a difficult gap to bridge without admitting some kind of a relation that is fundamentally external to the attitudes, and thus goes beyond Protagorean correctness. (I would say, beyond inferentialism as such, but this claim I have no space to defend here.)

Moreover, once we admit that some kind of an attitude-external relation is required to bridge the gap between language and the world, it seems that we need to change our understanding of the relation between the Protagorean criteria and correct applicability of a predicate. Namely, the incompleteness problem gestures towards the conclusion that the correct applicability criteria of some terms must be determined non-descriptively in some actual (i.e. non-ideal) context of application. So, the strict division of labour that formed a core part of the Protagorean account of correctness must break down somewhere. I tentatively argued that this somewhere is in the case of what Wright calls ‘basic concepts’, like colour and shape predicates. However, it can also be argued that many more complicated concepts also have their correct applicability determined independently of any descriptive criteria, as Kripke effectively did:

The example he [Putnam] gives is ‘cats are animals’. Cats might turn out to be automata, or strange demons (not his example) planted by a magician. Suppose they turned out to be a species of demons. Then on his view, and I think also my view, the inclination is to say, not that there turned out to be no cats, but that cats have turned out not to be animals as we originally supposed. The original concept of cat is: *that kind of thing*, where the kind can be identified by paradigmatic instances. It is not something picked out by any qualitative dictionary definition. (Kripke 1980; p. 122)

Although Kripke probably wouldn’t put it in this way, I think that his point is that sometimes, and especially with empirical concepts like ‘cat’, we should credit the world itself with ‘authority’ to determine their meaning (i.e. correct applicability). In fact, the authority should be seen as shared: it is the attitudes in combination with paradigmatic, actually perceived instances that together determine the correct applicability of ‘cat’. Of course, the world is not an authority in the same sense the attitudes are. But, then again, if the incompleteness problem stands, it seems that the attitudes aren’t quite as authoritative as that, either.

2 Conclusions

Peregrin’s primary concerns about meaning, it is fair to say, are about its normativity. The primary question he presents is something like ‘What makes a move in a language game

correct or incorrect?'. His primary answer, of course, are the rules of language, analogical to games like chess and football, instituted by normative attitudes.

The primary question of the externalist turn in philosophy of language, on the other hand, is not about normativity but *aboutness* of language. What makes a given word refer to whatever it does refer to? What is it for the word 'dog' to be 'about' dogs?

When all's said and done, I don't think Peregrin, as an avowed naturalist, should have fundamental problems negotiating the sovereignty of the attitudes, i.e. their ability to determine correct usage. In this regard, his version of inferentialism already starkly differs from Brandom's, whose main effort is to secure the scorekeeping practice as 'normative all the way down' (Brandom 1994; p. 625). Simple majority agreement even in the ideal limit of inquiry is not sufficient for Brandom's understanding of *genuine* normativity. In order to fully naturalise inferentialism, it must learn the central lessons of the externalist turn in the philosophy of language.

Funding Open access funding provided by Tampere University (including Tampere University Hospital).

Declarations

Conflict of interest No conflicts of interest are reported by the author.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright

holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Brandom R (1994) Making it Explicit: reasoning, representing, and discursive commitment. Harvard University Press, Cambridge MA
- Brandom R (2001) Articulating reasons: an introduction to Inferentialism. Harvard University Press, Cambridge MA
- Devitt M (1996) Coming to our senses: a naturalistic program for semantic localism. Cambridge University Press, Cambridge MA
- Devitt M, Sterelny K (1987)/1999 Language and Reality. Second Edition. Blackwell, Oxford
- Kripke S (1980) Naming and necessity. Harvard University Press, Cambridge MA
- McDowell J (1997) Brandom on representation and inference. *Philos Phenomenol Res.* <https://doi.org/10.2307/2953785>
- Peregrin J (2000) Meaning and structure: structuralism of (Post)Analytic philosophers. Ashgate, Aldershot
- Peregrin J (2014a) Inferentialism: Why Rules Matter. Palgrave MacMillan, London
- Peregrin J (2014b) Rules as the impetus of Cultural Evolution. *Topoi* 33:531–545. <https://doi.org/10.1007/s11245-013-9219-2>
- Peregrin J (2024) Normative species: how naturalized inferentialism explains us. Routledge, NY
- Raatikainen P (2020) Theories of reference: what was the question? In: Bianchi A (ed) Language and reality from a naturalistic perspective: themes from Michael Devitt. Springer, pp 69–103. <https://doi.org/10.1007/978-3-030-47641-0>
- Reinikainen J (2021) Brandom and the pragmatist quest for semantic objectivity. *Acta Philosophica Fennica* 97:55–78
- Rosen G (1997) Who makes the rules around Here? *Philos Phenomenol Res.* <https://doi.org/10.2307/2953786>
- Wittgenstein L (1958) Philosophical Investigations. Translated by Anscombe, G.E.M. Blackwell, Oxford
- Wright C (2001) Rails to infinity: essays on themes from Wittgenstein's philosophical investigations. Harvard University Press, Cambridge MA

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.