On Travelling Concepts

Martin Stokhof*

Proceedings of the Paris Institute for Advanced Study, 2025, Vol. 21

https://doi.org/10.5281/zenodo.15826893

Abstract
The paper discusses the idea of 'travelling concepts' in the context of 'philosophie pauvre', resulting in a Wittgenstein-inspired, pluralist but non-relativist view on conceptual structures. It is contrasted with that of various approaches in conceptual analysis and conceptual engineering. By way of illustration, the paper explores how a travelling concept view might help clarify discussions of understanding as applied to generative artificial intelligence systems.
Keywords
travelling concepts; 'philosophie pauvre'; conceptual analysis; conceptual engineering; understanding; generative artificial intelligence; Wittgenstein

Preliminary

This paper deals with a rather wide range of topics, each one of which probably deserves (at least) a monograph-length study of its own, and for each one of which there is an extensive literature. There is no way that one can do justice to all of that in the span of a single paper.

Now that may be a good reason not to try to do so, but rather to stick with one issue, one view. However, sometimes it can be useful to take a broad perspective, treat a variety of questions and observations as making up a single subject matter, one that can be approached from various angles. Sure, that does result in a lack of detail, but one may hope that one makes good for that by showing connections that otherwise would go unnoticed. This paper is written in that spirit.

---

* Department of Philosophy, Tsinghua University & ILLC/Department of Philosophy, Uni- versity of Amsterdam, m.j.b.stokhof@uva.nl, http://stokhof.org
This paper is the result of spending a month as writer-in-residence at the Institut d'études avancées in Paris. I would like to thank directors Saadi Lahlou and Paulius Yamin and the wonderful staff of the institute for their hospitality and support. Together with the stimulating environment created by the fellows of the institute, they made my stay not only productive but also extremely interesting and pleasant. I owe a special thanks to Saadi Lahlou for his support at a critical juncture.
I thank Michiel van Lambalgen for discussions on 'philosophie pauvre' and Tamara Dobler for introducing me to conceptual engineering and sharing her ideas with me. And thanks to Johan van Benthem, Tamara Dobler, Michiel van Lambalgen, Fenrong Liu, and Robert van Rooij for their helpful comments on an earlier version.

So, many details will be skipped, many interesting arguments will go unexamined and even go unmentioned. For a proper academic paper this is questionable. But we see no other way to present our take on the issues within the limitations that are set. So, we invite the reader to read this rather as an essay: a set of fairly general observations, questions and, yes, also arguments, that trace broad aspects of this problem complex and that we hope are of some interest. The paper is structured accordingly. The main text is devoted to a general and, we hope, accessible development of the main lines of thought, while references to and discussions of the literature are relegated to footnotes.[1]

That being said, what is the paper about? Concepts and their role in philosophy take centre stage. We start with an overview of key features of classical conceptual analysis, the dominant methodology in analytic philosophy (section 1). We continue with a discussion of the challenge that is proposed by conceptual engineering (section 2). Finding both lacking in certain respects, we sketch an alternative view on philosophical concerns, called 'philosophy pauvre', which comes with a different take on the nature of concepts, that of 'travelling concepts' (section 3). In order to illustrate the idea of a travelling concept, we discuss understanding as such a concept (section 4) and apply the results to some key issues in discussions of understanding in the context of generative artificial intelligence (section 5). We conclude with an outline of some general consequences of the views developed in the paper (section 6).

1 Concepts and conceptual analysis

Concepts: so central to philosophy, but also so confusing. Their nature: mental representations, platonic entities, behavioural patterns, linguistic expressions? Their function: trace independent features of real objects, project properties onto them, serve to organise experiences, guide behaviour? Everybody agrees that concepts play a crucial role in language and meaning, in knowledge and belief, in behaviour and values.[2] But what that role is and what concepts are that allow them to play that role: here we see the usual multiplicity of views that is so characteristic for philosophy. But before going into that in some more detail, it is useful to take a step back and briefly review how it has come to be that concepts occupy such a central position in the philosophical landscape.

If we take the long view on the history of philosophy, we see that it is not only a series of intellectual achievements, it is also a story of loss. Loss of territory, loss of epistemic authority, loss of method. Where in the battle for supremacy with theology philosophy arguably came out on top, it has subsequently lost terrain to physics, astronomy, biology, psychology, economics, sociology, anthropology, . . . Time and again philosophical reflection has given way to empirical investigation, often aided by the invention of new research tools and associated methodologies. Laboratory equipment replaced the armchair, observation and

experiment a priori reasoning, surveys and statistics introspection and thought experiment.

As a result, the status of philosophy as a source of knowledge changed as well. The rapid professionalisation of academic disciplines, the scaling up of academic research, the impact of research on economies, politics and other aspects of modern societies, tended to marginalise philosophy. However, although it was late to the game, philosophy has managed to carve out a space for itself in the academic landscape by following trends in other disciplines. Increased specialisation, both in teaching as well as in research projects and in publication venues, increased importance of publication records, a competitive labour market and more reliance on external resources, . . . In all these respects, philosophy has become an academic discipline like others.

The turn to conceptual analysis can be viewed as an integral part of the academisation of philosophy. For in order to carve out a position of its own in academia, philosophy had to do two things. First, it had to acknowledge that it had ceded authority on virtually all empirical topics to the various sciences. And second, it had to redefine itself vis à vis this fact. This is where conceptual analysis comes in.

One view, the roots of which we can discern in the early days of analytic philosophy, and which becomes dominant with the advent of logical positivism and its influence on Anglo-Saxon philosophy is the following. Philosophy no longer investigates empirical phenomena, but it does analyse the concepts with which the sciences do so. It becomes basically a second-order activity. The key here is that this form of conceptual analysis accepts a form of scientism: it is science and science only that delivers reliable knowledge. Philosophy reflects on the concepts that science uses to do that, it analyses them, and in some cases, it suggests clarifications or amendments of these concepts. But it is not in the business of providing knowledge.

There are other ways in which philosophers have embraced conceptual analysis, which are not exclusively science-focussed. The ordinary language philosophers in the second half of the twentieth century concentrated, not on the concepts that science uses, but on a different set of 'everyday concepts'. But that approach, too, refrains from making knowledge claims. Clarification, yes. Improvement in some way, perhaps. But knowledge, not so much.

Thus, concepts become the bread and butter of the working philosopher, it is what keeps them in business, alongside other academic disciplines. There are, of course, much more variants of conceptual analysis than that of the logical positivist or that of the ordinary language philosopher, but that does not need to concern us here. What matters is two things: that concepts have become the core topic of philosophy and analysis of concepts its main task, and that this state of things reflects a consciously or unconsciously held scientistic attitude.

So, all's well with the world and with philosophy? Not quite. For one thing, for the positivists their scientism limits what concepts they think are eligible for philosophical analysis. But why should other concepts be excluded?

For the more liberal ordinary language philosopher scientism does not impose a limitation on what they analyse, but it does limit the relevance of the results. Thus, we end up with a conception of philosophy as conceptual analysis that is limited.

Now that may be an acceptable consequence if one accepts the idea that science is indeed the only source of reliable knowledge. However, there is a second concern that is even more serious. It is this: concepts play a key role in science itself. There are two sides to this second concern. One is that reflecting on the central concepts that are used in a scientific field, or in a particular scientific theory, is not the privilege of the philosopher. Scientists themselves engage in reflection on and analysis of the concepts that they use as well. One could accept this and simply say: 'the more the merrier'. But where the concepts are closely tied to the way in which they are employed in actual scientific investigations, the results of the philosopher and those that are produced by the scientist may well differ. Should we then claim that those obtained by the philosopher are better, more relevant, than those of the scientist? This is not to doubt that philosophers are not good at conceptual analysis. They are (often). But it does raise the question what *special* expertise the philosopher has that makes their analyses of concepts better or more relevant than those of the scientists that work with these concepts.

And the second side is that concepts as such are of course a perfectly legitimate topic of empirical research. In fact, concepts and conceptual structures are heavily researched in cognitive psychology and cognitive neuroscience,[3] and there is research in social sciences and anthropology that is relevant as well. Again, we run into a demarcation issue. If the philosopher analyses concepts, then aren't the results of the investigations of, say, the cognitive neuroscientist relevant for that of the philosopher? If they are, isn't the philosopher then not engaged in empirical research after all? And if they are not, does that mean that philosophy has a domain of its own after all?

One answer to this demarcation question is provided by the view that the relationship between philosophical conceptual analysis and empirical investigation of the same concepts is asymmetric. This view bites the bullet and claims that philosophical analysis does not accompany scientific investigation, it precedes it: conceptual analysis is a prolegomenon to science. [4]

Philosophical analysis, the claim goes, is prior to and indispensable for scientific investigation because philosophy has a unique grasp on 'the bounds of sense', i.e., it can rule on what are proper concepts and what are not. That is not an empirical matter, but a philosophical one. So, given that the investigation of 'how things are' is an empirical one, it seems that such a view opposes realism with respect to concepts and conceptual structures. Apparently, it's not reality that decides 'what makes sense', because reality is the province of science. But then how is one supposed to evaluate the philosopher's proposals, the results of their analytic efforts? The connection between conceptual analysis and empirical investigation as activities may be clear, the former being prior to the latter. But

the matter of how one determines the adequacy of the former's results cannot then be an empirical matter. So, what is it? Does philosophical analysis result in a special kind of knowledge?

The upshot of all this is not only that philosophy has a domain of its own, but also that it has its own criteria to decide what is right and what is wrong. Philosophical expertise and what it applies to differ from scientific expertise and what science is all about. Philosophical conceptual analysis is distinct from both the empirical investigations and the conceptual considerations of empirical scientists. As a consequence, philosophical expertise is categorically, i.e., qualitatively different from scientific expertise.[5]

Quite another answer to the demarcation question that arises from the association of philosophy with conceptual analysis and the observation that science apparently has a stake in that as well, is not a hierarchical separation of philosophy from science, but rather the opposite: philosophy as a form of science.

On this view there is no specific philosophical domain, no specific philosophical method and philosophy delivers the same kind of knowledge as the sciences. Of course, differences exist, but they are quantitative, not qualitative. [6] This view avoids the questions about the differences between science and philosophy, and the special nature of philosophical conceptual analysis, that the 'prolegomenon'-view raises. On this take such differences do not exist and there is nothing special about philosophy. However, it does raise some other questions.

Also, on this view there still is something like 'philosophy'. But then how are its results and those of the sciences to be compared? Does it really make sense to say that 'philosophers are especially fond of abstract, general, necessary truths'[7] *without* investigating whether that is in fact true, and if so, why that is? Is philosophy more abstract than mathematics, or theoretical physics? That seems hard to argue.

Of course, once could claim that the difference lies in the familiar distinction between the necessary and the contingent. Now mathematics is certainly about the non-contingent, so the difference between it and philosophy is not covered by making this distinction. But one can indeed make the case that all the other sciences are not after the necessary in the way philosophy is. But isn't there then a principled distinction after all?[8]

If we want to maintain that there is a continuity between science and philosophy, this is not a possible position, it seems: philosophy as traditionally conceived is not a viable intellectual enterprise. But it what sense is it not much more radical and does it give up on philosophy as such? What is continuity between with science and philosophy other than some parts of science being done with greater emphasis on a particular type of question? Or perhaps, less rigorously than others?

Of course, there is a lot more to be said about this. But for our present concerns there is one conclusion and one question that are of central importance.

The conclusion is that in the view that science and philosophy are continuous one can discern traces of scientism.[9] It acknowledges that there are different ways in which scientific investigations are conducted, some more advanced than others, some more abstract than others. And some of these ways of doing science are called 'philosophy', presumably for traditional reasons. But in the end what delivers knowledge is science, and science only.

And a question remains: what about areas in philosophy that are not associated with questions of theory and explanation, but with practical and normative issues? The former are in the domain of science, at least as far as the empirical investigation of the relevant phenomena is concerned. But what about the latter? Obviously, modern philosophy has not completely abandoned their analysis. The traditional positivistic argument that what look like normative issues are either empirical ones in disguise, or nonsensical to begin with, has failed.[10] Conceptual analysis is a key part of the way in which philosophy addresses question about value, norms, aesthetics, and so on. This has manifested itself in a distinct shift towards 'meta-level' analysis. Ethics has given way to 'meta-ethics', in political and social philosophy emphasis has shifted to analysis and clarification of relative conceptual complexes, and in aesthetics the meta-theoretical perspective also dominates.

Quite generally, philosophy has withdrawn from hands-on engagement with subject matters (the good life, justice, art) and has focusses on the analysis of conceptual structures, adequacy criteria, notions of explanation and justifications that shape the scientific theories that do deal with these subject matters directly. As a consequence, it has conceded authority to science and severed a good number of the many links it had with problems and questions that really matter to us. It has turned philosophy into a highly specialised discipline, one which is, perhaps, valued for its adherence to high standards of intellectual rigour, but which also has lost much of its relevance for some of the key issues.[11]

But there have also been movements in a different direction, in which practical concerns maintain a central position. One of those is conceptual engineering, the other 'philosophie pauvre'. They form the subject of the next two sections.

2 Conceptual engineering

Like so often in philosophy, the singular term 'conceptual engineering' covers a variety of approaches and views. And indeed, on the 'inside' differences are emphasised and arguments for and against certain positions are given a lot of weight. For an outsider, however, there are enough commonalities to continue to use the singular term, although the differences between two broad trends are noteworthy, as we hope to show later.

Two interrelated convictions inform conceptual engineering. The first one is that there are good concepts, i.e., concepts that promote worthy moral or political goals, and bad ones, that hinder the realisation of those goals. And the second one is that the task of philosophy does not stop at analysing existing concepts, but that it should also be involved in designing ('engineering') better ones.

Many topics that conceptual engineering has been concerned with come from social and political philosophy and ethics, for example, concepts pertaining to gender and race.[12]But also outside the sphere of practical concerns there have been attempts to conceptually engineer concepts such as truth and belief.[13]

It is clear that, no matter the particular (set of) concept(s) that is at stake, conceptual engineering is a normative enterprise.[14] Conceptual engineering is not just a matter of increased descriptive/explanatory adequacy: that kind of conceptual adjustment is bread and butter of scientific development. Rather the idea is that by prescribing new concepts we can pro-actively induce change. Concretely: change that we want, change that (we think) is needed for (broadly conceived) reasons of social and political justice.

That immediately raises some questions. What kind of normativity are we concerned with here? It is clear that conceptual engineering assumes that normative assessment of concepts is not arbitrary but somehow objective. Does that mean that it is assumed to be based in reality? If so, what kind of reality? In what is considered as such at some point in time in some community? Or in some form of reality that transcends contingent historical and social determinations?

And with the question to the kind of normativity comes the question of who has access to it, who has normative authority. Conceptual engineering being conceived of as, at least also,[15] a philosophical endeavour, does that ascribe moral authority to philosophers? And if so, to philosophers only? In view of the fact that many conceptual changes are brought about by science, not by philosophy, it seems awkward to view conceptual engineering as an exclusively philosophical affair.

Another set of questions raised by the goal of conceptual engineering to induce change concerns effectiveness. The idea is that conceptual engineers design better concepts, or, even more ambitious, the right concepts. But that's only one part of the story, there also has to be a follow-up. How are these better, or best, concepts going to be put into use? For example, if we observe that the discussion about a certain normative issue, say one concerning gender and equality, is marred by (some? all?) discussants using the wrong concepts, who is going to persuade them to use the better ones? And how? Under what conditions are these attempts going to be successful?[16]

Obviously, conceptual engineering as a form of philosophy strives to be a practical enterprise, not a theoretical-analytical or an empirical one. In that respect, its goals differ from those of the various forms of conceptual analysis that we outlined in section 1. These analyse what there is, conceptual engineering is (also) concerned with what there should be. In light of that, it

is interesting to note that some of its proponents see conceptual engineering as a natural continuation of conceptual analysis. We will go into that in a bit more detail below. At this point, we note that this might indicate that despite their differences, conceptual analysis and conceptual engineering nevertheless do share something, viz., a scientistic outlook. After all, the term 'engineering' will not have been chosen lightly, and it suggests that in the same way that engineers build on and use the results of science, conceptual engineers build on the results of conceptual analysts, and, indirectly, on those of the science that uses them.[17]

At this juncture, it is useful to briefly discuss two main views in conceptual engineering that have been developed over the years. These differ in how they see the interactions between conceptual engineering as a philosophical undertaking and research in social sciences and humanities. We can call them the 'thick' and the 'thin' view.

The 'thick' view is the one that is held by those who are concerned with conceptual engineering of concepts that play a role with concrete topics and who actively work with the results of research in the sciences on those topics in order to effect actual changes.[18] The 'thin' view would be endorsed by scholars who take a more abstract stance and, by and large, drop the goal of inducing actual changes.[19] One way of fleshing out the difference between the thick and the thin view is by looking at how each looks at normativity and its grounding.

Thick conceptual engineering grounds normativity in the world in an objective way, and it is precisely for that reason that it can align its endeavours with those who study the issues from an empirical angle. Adherents of the thin view suggest that this makes the thick view collapse in some form of descriptivism, which contravenes their much more abstract approach. But is that a fair characterisation? For one thing, we should note that 'grounding in the world' can mean quite different things. For example, if we think of 'the world' as a socially structured whole of practices, communities, and cultures, we are dealing with a plurality, and grounding a particular construction of a concept in one specific part of that plurality does not amount to pure descriptivism. It involves making a choice, singling out one way of doing things from a plurality of such ways.

Note that this characterisation assumes that the preferred construction is already there, i.e., it is already manifested in a particular way of doing things. Of course, that is not always the case. So we are led to a broader conception of 'the world', one in which we can ground the result of our conceptual engineering: it includes not just actual ways of doing things but also imaginary ones.[20] We think up a way of doing things that grounds our engineered concept. Now a different question arises, which has to do with realisability. It is one thing to think up how things should be done, but what we want is to think up a way that is actually realisable, one that, although it is imaginary and as such not actual, we can in fact actualise.

But we can construct 'grounding in the world', of course, also in a

metaphysical way, by claiming that our engineered concept corresponds better (or, ideally, completely) with 'the way things objectively are'.[21] If we do that, then, given that we are dealing with concepts that have a substantial normative dimension, it seems we are committing ourselves to some form of moral realism. It is not quite clear why such a commitment to moral realism constitutes a collapse into an objectionable kind of descriptivism. But the commitment itself is something that one may want to avoid.

As for the thin form of conceptual engineering, that seems to lack sufficient constraints to prevent it from collapsing, not in some form of descriptivism, but rather into a form of relativism[22]. And the question then becomes whether such a thin, relativistic conception of conceptual engineering makes sense to begin with. Can there be conceptual engineering without a proper basis of normative judgements? One defence against this objection could be that we normally have, and, if we do not, should have, normative judgements about concepts.[23] However, unless there is an *independent* grounding of these judgements, i.e., one that is not induced by our conceptual engineering efforts, there seems to be no way to prevent the thin form from collapsing into relativism.[24] Now, we do not need to take up a position in the debate between thin and thick conceptions of conceptual engineering. The only relevant takeaway for our purposes is that the very existence of the debate shows that the role of normativity is key. Ideally, one would want a view on how concepts function and change, and in some cases can be made to change, that allows us to do two things. First, to avoid the relativism that threatens the thin view on conceptual engineering, and, second, to avoid the commitment to some form of metaphys- ical grounding that the thick view seems to require.[25] As we hope to show in the next section, the 'philosophie pauvre' conception of philosophy and the notion of travelling concepts that it harbours provide a framework that steers clear of these two problems.

3 'Philosophie pauvre' and travelling concepts

Not conceptual analysis, either in the traditional way or as a prolegomenon. Not philosophy as continuous with science. Not conceptual engineering, either thin or thick. What possible conception of philosophy is left?

What we aim to show in this section is that the considerations brought forward in the previous sections not only tell us what philosophy is not, they also trace the contours of what it could be. Philosophy is not beholden to merely analysing the concepts of science. It need not, in fact, it cannot, claim that its analyses have priority over empirical investigations. It does not need to hold that its investigations are akin to those of science. And it does not need to claim to be the authoritative source of good concepts, whatever that may be.

The main reason that makes philosophers claim that philosophy is any of these things is that, knowingly or unknowingly, they define philosophy in relation to science. The dominance of science explains that, but it does not make

it unavoidable. We can conceive of philosophy also as a pursuit of a different type of understanding. In that pursuit, it may engage with the same phenomena as science does, but it does so with a different aim.

This puts philosophy on its own footing. Not because it has a domain of its own, or its own special method, but because it looks at things in a different way.[26] Where science investigates by observation and experiment and aims at an understanding that is explanatory in nature, philosophy looks at things in terms of the meaning they have for us: how things affect our actions, our understanding of the world and of ourselves, the practical significance they have for us in our lives. It does not deliver the kind of propositional knowledge that explains things in terms of causal laws. It is rather a kind of know-how, a way of looking at things that reveals the meanings they have and might have for us.

This is a philosophy that is not modelled on science with the limitations that come from that, but that is also not restricted to the purely therapeutic enterprise of just clearing away conceptual confusions. It is one that is substantial, i.e. that deals with real phenomena, that is co-operative, i.e., that makes no claim to being able to deliver special, 'one-of-a-kind' insights, and that is modest, i.e. that makes no claims to any special authority or intellectual priority. It is a 'philosophie pauvre'.[27]

One way to frame this conception of philosophy is in terms of Wittgensteinian 'aspect seeing'. We cannot go deeply into the details of that here,[28] but it is useful to highlight a few key elements. 'Aspect seeing'[29] is about a particular way of changing perspectives. It is not about seeing things differently in the sense of seeing them in the right way, instead of wrongly. It is about broadening the range of ways in which we can see things. The familiar example of visual aspect seeing, the duck-rabbit case, provides a nice illustration. At first, we may see the picture only as a duck. After the aspect switch has been realised, we can also see it as a rabbit. But clearly, there is no right and wrong here. It is just that we now have two ways of seeing the picture, instead of one. When we see the duck-rabbit as a rabbit, we are not seeing it better than when we see it as a duck. And it is also not about getting rid of a way of seeing. You cannot *not* be able to see the duck-rabbit as a duck after also having become able to see it as a rabbit. It is about seeing things not just like this, but *also* like that. Aspect seeing is about freedom.

Philosophie pauvre is about bringing about the same kind of change, it is an aspect seeing, aspect switching endeavour. What it aims to do is to bring about a change in the way we see things, in the way we think about them, and therewith in the meaning they have for us. The nature of that change is crucial. As we already noted, it is emphatically not about replacing a 'wrong' way of seeing by a 'better', let alone the 'right', way. [30]

What philosophy does is create an openness, a space of possibilities for us to explore. And thereby it can free us from an obstinate, 'compulsive', one-sided way of thinking about a certain issue.[31] That space is not determined by

reality, so changing our ways of seeing is not guided by ontology. If anything, the change is 'epistemological': it affects how we look at the world. But that, too, is not fixed: the space of possibilities itself can change over time, according to changes in our needs, changes in our means of interacting with the world.

Our abilities of aspect seeing and aspect change reflect our awareness of the intrinsic plurality of our engagement with the world (including ourselves). That ability and that plurality is not intrinsically philosophical. It occurs, and is needed, across the board: in everyday life, in the arts, in science, . . . But it *is* the perfect antidote to philosophy's chasing necessary, universal truths. [32] Creating this freedom is not restricted to a particular kind of issues,[33] nor is it determined by pre-defined goals and methods. In that sense neither domain nor method is what defines 'philosophie pauvre'.

So, philosophy creates freedom, but that freedom is not without limits. Freedom is an effect of philosophical observation, but it is productive only because it is subject to limitations. Unlimited freedom is meaningless since meaningfulness requires change *and* stability. And as a matter of fact, the stability of the world is a source of limitations of our conceptual freedom. These are a productive kind of limitations, they create the space within which different ways of seeing and imagining things can be identified and explored. One could say that without the limitations the differences would not make sense. Meanings exist because they range over different situations that are nevertheless comparable in certain regards. I.e., meanings allow us to identify what is the same (stability) across different situations (change). [34]

But change and stability are not enough, there also needs to be relevance. Philosophy creates meaningful freedom only when tied to practical, everyday concerns, to what makes sense for us to do, to what is required by the concerns of everyday life, including moral concerns. This is reflected in the fact that meanings reside in *practice*s, ways of acting and interacting, verbally and non-verbally, that 'have a point'.

Practices are constituted by certainties, i.e., by convictions and ways of doing things that are beyond our usual procedures of justification.[35] Certainties, and hence also the practices that they make possible, are not forced upon us by reality. Different communities have different ways of doing things, our practices change over time, we can actively change them, and even if we cannot, we can imagine them to be different. So, there is freedom here as well, but again, it is limited. A practice has a point: it is a conglomerate of actions and interactions that serves a purpose. Educational practices, building practices, farming practices, artistic practices, . . . they form a manifold of great diversity, but they all have one thing in common: they are there for a reason, people engage in them in order to get something done.[36] That can usually be realised in a variety of ways, the pluralism of ways in which practices play out testifies to that. But this pluralism is not a relativism, not anything goes: nature imposes limitations.

'Nature' is used here in a broad sense. It refers to the physical environment in which we live, its laws and characteristics that help shape the space of possible practices that we can develop and maintain. It also refers to ourselves: the way in which we are embodied, the needs that arise from that, our perceptual capacities: a wide range of features of what we are as a biological species that likewise enables and limits. And we can include basic human psychology in the mix as well. It is the ways in which we as living creatures are attuned to our physical niche that make a range of possible practices available to us but also exclude some that might be available to other types of animals, or that would be available to us in a physical environment with different characteristics.

So, our space of possible practices is limited but not determined by nature. Some of the limitations are stable, others less so. The basic laws governing our physical environment are of the first kind. They can be imagined to be different (well, some of them, at least), and for sure our knowledge of them changes over time, but as far as our practices are concerned, they are what they are. That being said, it is also true that our actual physical environment represents just one possibility that these basic laws allow. And, for better or worse (mostly for worse), we are changing that environment in ways that also have an effect on our practices. And the same goes for ourselves: there are certain characteristics of our biological and psychological nature that seem fixed. But others can be changed and are being changed. We change our bodies with artificial limbs or with non-human organs, we extend our perceptual range of possibilities with a wide variety of instruments, we enhance access to and processing of information with computational tools, and so on. Thus, nature in the comprehensive sense of that which defines the space of practices available to us is itself both stable and dynamic.

Philosophical reflection is one way we have of becoming aware of this intricate network of limitations, possibilities, changes. Not the only way, of course: the very same phenomena are studied in a variety of ways, in the historical sciences, in anthropology, in psychology and human biology, for example. Some focus on aspects of change, others more on the stable elements. In general, they aim for explanations, in terms of law-like generalities. Philosophy adds to that, but as a companion, not as a rival. Its focus is more on how core concepts are used and thus have meaning across a variety of practices.

Humans live in the same natural world and have the same human nature. But those communalities go hand in hand with differences. Different communities (historical, contemporaneous) have different knowledge of, and different views on that nature. And within a community there are different practices, different ways of engaging with nature, ranging from science, to art, to the everyday. The ability to see and think differently is of crucial importance, and this is where philosophy as critical reflection has a role to play. It allows us to trace the constitutive elements of our world picture, both the natural ones as well as those that are community-specific. It does so by reflecting on our practices, by coming to see the aspectual nature of some of what constitutes

those practices, by seeing and investigating new aspects, and by thus creating tentative alternatives.

Imagination has a key role to play here and it is imagination that makes this kind of investigation different from empirical ones.[37] At any given state of the external constraints we may be able to 'think outside the box' that they constitute, i.e., think of concepts that are not actually candidates for adoption, because they do not answer to the constraints. We can imagine such concepts, but we cannot adopt them because they do not work given the constraints that actually obtain. But they might work if the constraints were different.

Stability and change are also reflected in the way we use concepts. Our need for stability is reflected in the fact that we often use the same concepts across our various practices. But that does not make those uses the same across these practices. Example. In the context of an everyday conversation about why someone is mad at somebody else, the concept of explanation is used in a different way from how it is used in the context of a discussion about the results of running some experiment, or in that of teaching calculus to high school students. Does it make sense to say that one of these is 'the right' concept of explanation and that the others are somehow lacking? That would seem a very strange reaction. There is no right or wrong here, no such thing as 'the real concept' of explanation. Each way we use it in context is attuned to what is relevant in that context.

That is *not* to say that every use that someone makes of a concept is right. Obviously, a concept can be used in the wrong way. And there can certainly also be situations in which the way a concept is used is in need of change. Whenever we feel we need more precision we make the necessary arrangements. We do so in science, for example, when new findings change our insights into how events are causally connected, which are then reflected in how certain concepts are applied, in the context of explaining the events. We do so in legal contexts, for example in a contract where we start with a set of explicit stipulations how certain terms are to be understood in the context of that contract. And we do so time and again in everyday conversations. ('Let me get this straight. So, you understand by . . . ?') However, and this is the crucial point, changing a concept is not *directly forced* on us by independent considerations, of an ontological or axiological nature. Given that there are always alternatives, ultimately it is a matter of decision. This goes against the standard view that concepts, at least the good, 'right', successful ones, are rooted in real, ontological distinctions, and that it is that independent ontological structure that determines whether *we* have a proper grasp of what reality is like, i.e., whether *our* conceptual structure are 'true to the facts'.[38]

This varying with practices is what we call the travelling nature of concepts, and concepts that display this nature we call 'travelling concepts'.[39] Our use of the term aims to capture the fact that we use many concepts in a wide variety of practices: in science in its many forms, in politics, in legal practices, in commerce, in everyday conversations, in the arts, and so on. The same lin-

guistic expression is used, and in an important sense that means that the same concept is used. But the concept takes on slightly different features depending on the specific context in which it is used. Thus, it 'travels' from context to context, recognisable as the same, but sufficiently different to suit different circumstances, to be part of different practices.[40]

This dependence on the multiplicity and variety of practices distinguishes the present notion of travelling concepts from the kind of conceptual change in development of scientific theories, that is studied in the history and philosophy of science, such as Kuhnian paradigm shifts. And it also differs from the notion of travelling ideas, that is a central notion in the study of intellectual history and the history of ideas. Paradigm shifts and travelling ideas are concerned with *changes over time within the same practice*. Travelling concepts, on the other hand, are concerned with *commonalities and differences across different simultaneous practices*. The simultaneity they share with the Wittgensteinian notion of family concepts,[41] but they differ from the latter in that the differences arise from different practical contexts.

The analysis of concepts as travelling concepts also differs from traditional conceptual analysis and from conceptual engineering. Each in their own way the latter are committed to the assumption that there is objectivity when it comes to concepts, and in the case of conceptual engineering that this objectivity has moral authority. Philosophie pauvre and its notion of travelling concepts also has a moral dimension, and in that respect it 'sides' with conceptual engineering against the scientistic mentality of traditional conceptual engineering. But this moral dimension is of a different nature. Conceptual engineering works with a conception of correctness that would allow us to motivate *replacing* one set of concepts by another, more correct one. But the changes that philosopie pauvre is after are never about replacement, they are about getting rid of the idea of there being one, ultimately correct set of concepts, and any claim to authority that philosophy might stake out, is rejected. At the same time, it acknowledges the limitations that nature, broadly conceived, imposes. It is in this sense that it embodies a 'pluralism without relativism'.[42]

This allows philosophie pauvre to stay clear of a principled objection to conceptual engineering, viz., that it turns our concepts into 'mere' social constructs.[43] Here pluralism without relativism works: the acknowledgement that our conceptual structures are constrained by nature (in a broad sense) prevents the collapse of a non-metaphysical account into a kind of relativistic social constructionism without thereby committing us to a kind of metaphysical grounding.

Of crucial importance is that these external constraints are not groundings, in the sense of justifications. They are de facto constraints that constitute a sphere of possible conceptual structures. But this constitution does not allow for a definition of what the possible concepts are, since the external constraints themselves are contingent. Nature may change. In fact, it does change and in

some cases we are the ones who are changing it, deliberately in some cases, unintentionally in others.

This also explains the role of imagination. At any given state of the external constraints, we may be able to 'think outside the box' that they constitute, i.e., think of concepts that are not actually candidates for adoption, because they do not answer to the constraints. We can imagine such concepts, but we cannot adopt them because they do not work given that constraints that actually obtain. But they might work if the constraints were different.[44]

One question that arises is how a philosopher who makes free use of their imaginative capabilities with regard to ways in which concepts can be used, can stay clear from essentialism? That we are able to imagine different practices than we have, is obvious, and it is equally obvious that our imaginative capabilities are a crucial factor in almost everything that we do. Imagination is not just key to literature and other aesthetic practices, it also plays a central role in science, and it is an essential ingredient of our everyday lives: planning and decision making depend on it. The question then is this: do the limits of our imaginative capabilities with regard to how a concept can be used coincide with what we would like to call the 'essence' of that concept?

At this juncture the crucial observation is this, that our imagination itself is a 'moving target' in the sense that it, too, is determined by historical, social, cultural, . . . factors. If that is correct, then philosophical analysis in Wittgenstein's sense, in which imagination plays a key role, does indeed occupy a position that differs both from that of scientific inquiry, which is factual in nature, and from that of traditional philosophical analysis, which aims at uncovering essences. This is the difference between an investigation into 'What is *X*?' (factual, as in science, or essentialist, as in traditional philosophy) and an investigation into 'What is *X*-for-us?'. The concepts that philosophie pauvre deals with are 'travelling concepts' and its method, too, has a 'travelling' dimension.

Another question that arises then is this: what determines what are possible travelling concepts? One observation concerns the role played by 'facts of nature'. As we have seen above, these provide limits, by do not force a particular set of concepts on us, they determine a space of possibilities, different sets of concepts that we can employ practically. Now what is interesting to note is that, on the one hand, these limits set by nature are a given: reality (nature) is what it is, but that, on the other hand, we can imagine them to be different (think: science fiction) but not in a completely unlimited way. So, there is a complex dynamics between what is the case and what we can think differently, and that is a dynamic that itself changes also.

That might suggest a very open space of speculative possibilities and possibilities to speculate, but as far as Wittgenstein is concerned there is third factor that plays a key role and that limits what makes sense to do with and in that space. It is the matter of 'having a point': 'The game, one would like to say, has not only rules but also a point' (*Philosophical Investigations, 564*).

A language game or practice needs to have practical value, and that applies both to the ones that we have as well as to the ones we can imagine. Now practical value itself is a very diverse concept. However, it does restrict what makes senses to do in this space of imaginative possibilities. Pure imaginability is not enough: any practice, be it factual or imaginable, needs to have practical value, practical meaning for us to be a practice to begin with.

This gives application in a practical sense ('having a point') a central role and, consequently, the notion of a travelling concept is not strictly analytic and strictly propositional. The general idea is that of a concept that gets a different content, and hence a different application, depending on the context in which it is employed. One particular type of context is that of 'investigation' (analysis) (scientific, philosophical, . . .): depending on the kind of investigation (goals, methods, . . .) a concept may be employed in different ways, and produce different results. An interesting case, is where the results of the investigation change the concept that gave rise to it.[45] Arguably, some of the central concepts that play a role in our understanding of ourselves ('mind', 'language', 'meaning', 'reason', . . .) are of this kind. We employ these in our self-understanding, we also investigate how they are employed, and the results of that investigation may change them. Note that the assumption that there is a fixed reality that these concepts correspond with is at odds with this view only on the assumption that this reality is 'cut at its joints' by the concepts at some particular point in time. But if we allow for this reality to be fluid in the sense of allowing for different conceptualisations, which are not internally strictly ordered by a relation of accuracy, then we can acknowledge the travelling, or looping, nature of these concepts and at the same time maintain that they correspond to a reality beyond them.

Philosophie pauvre, as the kind of philosophy that embodies this view, is intimately tied to what we are and what we do, but also to what we can be and what we should do. As such is not a purely intellectual activity, but an *existential* one, and being existential, it also has a normative dimension. First and foremost, it is an attitude, not a separate set of problems, not a distinctive method, but the exploration of ways of looking at things from the perspective of practical concerns, of 'what matters to us'. It acknowledges the limitations of science, and of the particular type of rationality it embodies, *without questioning it in its own sphere*. It accepts that outside the sphere of science, we need to be modest. Things are outside the sphere of science for a reason. They are difficult but cannot be decided in a factual manner: reality does not dictate our concepts, but it is also not arbitrary which concepts we have and how they change, because there are often important consequences that are (also) normative in nature. Therefore, it emphasises the need to be modest in our claims, always with a willingness to see things differently, to accept the lack of an ontologically grounded objectivity. As a consequence, it also refrains from absolute moral prescriptions and remains (self)critical without end. And that by itself is taking a moral stance.[46]

4 Understanding as a travelling concept

In the remainder of this paper we will be looking the concept of understanding as it is applied in discussions about generative artificial intelligence and concrete systems based on that.[47] Our starting point will be an analysis of understanding as a travelling concept. It is loosely based on Wittgenstein's view of understanding as an ability.[48] In this section we will introduce that analysis and examine some of its key characteristics. In the next section we will then apply it to discussion about understanding as applied to genAI and genAI systems.

In *Philosophical Investigations* Wittgenstein discusses understanding in the context of what are called 'the rule following considerations'.[49] Rules and concepts and meanings are closely related. Concepts are associated with meaningful expressions and rules pertain to their application in language games. Such rules need to be learned, and thus taught, and understanding kicks in as a mark of success. When we teach someone how to follow the rule for the application of a certain concept/expression, we need to be able to indicate when we think our teaching is successful, i.e., when someone has understood.

The rule following considerations investigate the epistemic conditions under which we ascribe, or deny, understanding. Various interpretations of Wittgenstein's arguments have been proposed but that not need not concern us here.[50] Our main take away from Wittgenstein's scattered, and often somewhat enigmatic, remarks is that ascription of understanding is first and foremost ascription of an *ability*.[51] Someone has understood if they are able to do something in the right way. Now that is a very general characterisation that covers different kinds of cases. A full phenomenology of understanding is beyond the scope of this paper, but the following observations will be important.

A first observation concerns the role of language. There are obviously situations where the ability that manifests understanding is a verbal one, where being able to use a linguistic expression correctly serves as the criterion for understanding. But understanding can also manifest itself non-verbally, e.g., in being able to correctly sort objects using some feature as a way of distinguishing them from one another. In those cases, the understanding is not linked to a linguistic entity and its use, but to the grasp of some property of objects and its use as a distinguishing feature. So understanding is not necessary something tied to language.[52]

Secondly, if we view understanding as an ability to do something, in a broad sense of 'doing', we also de-emphasise the role of the mental. For sure, with understanding comes accompanying mental states and processes, that should be obvious. But on the ability view there is no *identification* of understanding with some specific mental state or process. Understanding is not reduced to having a mental representation or going through some mental process.[53] Rather, it is viewed from a functional perspective. If to understand is to be able to act, then anything that has the same action potential can be ascribed, or denied, understanding.

Usually, functionalism is associated with 'liberating' some concept from a particular form of material constitution, a specific material substrate. Take representations. Functionalism holds that these can be materially realised in a wide variety of material substrates, ranging from configurations of wet matter to written and spoken language to assembler code running on silicon. That makes it possible, for example, to hold that both humans and computers have representations, meaning that they are in states that are materially different but that nevertheless function in the same way. Inasmuch as the concept of understanding as ability likewise dissociates understanding from the mental and the linguistic is has a similar 'liberating' effect.

However, it is of crucial importance to note that if we view understanding as an ability to act, we *also* impose restrictions: whatever we ascribe understanding to must have a sufficiently similar action potential. When it comes to humans their action potential is closely tied to their embodiment, to the kind of bodies they have and the actions those bodies allow them to perform. It is important to get this right. It is not the case that our bodies *define* our action potential. We have developed ways to extend the action potential of our physical bodies in various ways. Tools allow us to act upon things with more force, from a distance (physical, temporal), with greater accuracy, to observe with increased resolution, and so on and so forth. The development of language is perhaps the most important step in human evolution, allowing us to pool resources and act collectively in complex ways, to influence and manipulate each other without being confined to using our bodies. Of course, language in many ways remains also a bodily activity, but what we can achieve with it escapes those limitations.

However, and this is key also when it comes to understanding and genAI, our embodiment does shape even our extended action potential in that it determines *what makes sense for us to do*. Recall the perspective that was introduced in section 3. Practices, we emphasised there, are coherent conglomerates of verbal and non-verbal actions *that have a point*. And possible practices are *constrained by nature*.

Now, having a point is very much linked to our needs and our abilities, and these are shaped by the way(s) in which we are embodied. And the constraints of nature include those imposed by our nature, and that in its turn is shaped by our embodiment as well.[54] So both in what possible practices we might have as well as in what actual practices we do have, our embodiment plays an important role. And that is reflected in the concepts that characterise our practices. They are not strictly defined, nor are they strictly definable. They are travelling concepts. But the ways in which they change from context to context, their possible itineraries, are also shaped by the kind of creatures we are, including by the way(s) in which we are embodied.

So, when the concept of understanding, being a travelling concept, is applied in new contexts, to new kinds of entities, it will have a natural tendency to retain features that derive from its original contexts. Whether this tendency

constitutes a limit, or whether it can be set aside by other concerns, will be a key question when we look at understanding as applied to genAI systems. More on that in section 5.

Before going into that we need to look at some further characteristics of understanding as a travelling concept. Where do we see its context-dependence manifested? One manifestation of its travelling nature is displayed by the context-dependence of the linguistic expressions that we use for ascribing understanding.[55] We stick with 'understand' and 'understanding'.[56]

Both terms are gradable and relative. A gradable expression is one that allows gradations. For example, 'be annoying', or 'trust'. We can use them with qualifications such as 'very', 'a little', 'somewhat', 'completely'. Someone's behaviour can be said to be 'very annoying', or 'a little disappointing. We can 'trust someone completely', or find them 'somewhat awkward'. The grading expressions typically invoke a scale.

Examples of relative expressions are adjectives such as 'tall', 'expensive'. We say of a 12 year old that they are 'tall for their age', or of a basketball player that they are 'tall for a point guard'. But obviously 'being tall for a 12 year old' is not the same as 'being tall for a point guard'. Relative expressions typically invoke a comparison class. Note that relative expressions are usually gradable: 'very tall for twelve-year-old', 'too expensive for a daily wear', but not all gradable expressions invoke comparison classes: '? awkward for a . . . '

What about 'to understand' and 'understanding'? These may note that these are gradable expressions. We say that someone 'understands something completely / to some extent / . . . /not at all'; that they 'have a complete/ insufficient / . . . understanding of' or display 'a complete lack of understanding'. Interestingly, 'to understand' and 'understanding' are also relative expressions. They invoke comparison classes, as is shown when we say of someone that they 'understand well (enough) for a four-year-old/high school kid / lay-person / . . . ', or that they 'display expert understanding / complete mastery /… ' [57] Reference to a comparison class is not to be confused with invoking a property as an explanation, as when we say of someone that 'being an expert, they understand how this machine works'.

That 'to understand' and 'understanding' are gradable and relative expressions reflects the travelling nature of the concept of understanding. Their gradability is not directly due to context-dependence, but it does highlight that understanding is not a yes/no affair. It comes in degrees, so it involves the application of a normative criterion: this is insufficient understanding, that is proper understanding. It is the invocation of a norm that allows us to make fine-grained distinctions. The invocation of a norm introduces a parameter that can take on different values and that opens up room for a more principled form of context-dependence which is reflected in 'to understand' and 'understanding' being also relative expressions. Their use invokes a comparison class: they are used to make true or false statements only relative to such a comparison class. And in that they reflect the travelling nature of the understanding concept.

Another feature of understanding that reflects its travelling nature is the connection with the autonomy of the behavioural responses it gives rise to. Understanding a situation or an event is typically displayed in various ways of responding to that situation or event. We see the raindrops on the window pane, understand that it is raining and grab an umbrella when leaving the house. We calculate how much force it takes for a specific type of rocket to reach escape velocity; we determine how much fuel is needed; and then we load that into the rocket. We understand that someone has been hurt by a careless remark we made and apologise or make amends in some other way. Some cases are complex but pretty deterministic: the rocket fuel case. Others are matters of habit: the umbrella. And yet others are perhaps simple, but also much more open: the careless remark.

What is important is that understanding and responding display both context-sensitivity (what is the specific, concrete context?) as well as autonomy (what is the larger, long(er)-term context?). Example: my watch is set to remind me to stand every hour. That is a very specific type of context-sensitivity that gives rise to a deterministic response. But my watch does not know that I am ill and that hence I shouldn't be prompted. It lacks relevant information, and it does not have the ability to determine the proper response in different situations. In that sense the watch does not understand: its response lacks autonomy.

This illustrates a general point, viz., that understanding is embedded in a web of abilities. It is connected with responding, caring, deliberating, deciding, and so on. When we understand, we display our understanding in taking certain actions (or in refraining from them). That means that we make choices, predict and evaluate effects, and so on. And we do all this in a way that is closely tied to the specifics of the situation at hand. We do not ascribe (or deny) to someone understanding in isolation, we base such judgments on a whole range of relevant actions, verbal and non-verbal.

This has important consequences. For one, it indicates that it does not make much sense to try to identify understanding with some structural condition of whatever it is we that ascribe understanding to, such as having a particular type of representation of a situation or event or making certain types of inferences about that. Of course that is involved, some way or other. And in a strong way: 'Ohne Phosphor kein Gedanke!'[58] But 'is involved' is not 'is identical': understanding cannot be reduced to the representations and inferences that it involves. And although there is always some physical substrate, 'some way or other' leaves room for a variety of them.[59]

Another consequence, one that will be important in what follows when we look at understanding in the context of genAI systems, is that due to its connections with other abilities, understanding is multi-dimensional. Not only can it not be tied with a particular type of substrate, it also cannot be identified with a particular type of functionality. Understanding is displayed in many different ways and is ascribed (or denied) on the basis of different types of evidence.[60]

5 Understanding and generative artificial intelligence

It is no exaggeration to say that recent developments in genAI and the advent of ever more capable genAI systems[61] have shaken up the way we look at what these technologies can do and will be able to do in the near future. Where the very idea of an artificial general intelligence[62] that would be able to compete with humans, and surpass them, for a long time was a matter of science fiction, it has now projected itself as real possibility, something that will affect us in complex, profound, and largely unforeseeable ways.

We use the term 'project' on purpose, because there still is a definite disconnect between the idea of an AGI and the actual capabilities of working genAI systems. In the mind of some that is a gap that will be closed in the coming two or three years, others think it will take much longer. Certainly, there is rapid development in particular areas. And, although there are some that dismiss the very idea of an AGI as such, there are not that many who think that the project as such is illusory.[63]

One of the reasons that opinions diverge is that discussions about the present and future capabilities of genAI systems are conducted in a variety of registers, which makes it very hard to determine which views exactly are being defended or challenged, and with what arguments the discussions are being held. To give one simple example: there is no accepted definition of what an AGI is or should be. 'Something that is smarter than humans'. That sounds nice, or frightening, or both. But what does it mean? What do we mean with 'being smart'? And how smart are humans? Is that how smart the smartest is? the dumbest? 'the average human'? And smart at what? solving math problems? coding? tying shoelaces? dealing with ethical dilemmas?[64] telling bedtime stories? consoling a sick child? The confusion is profound, and we are not even going to try to clear things up.

But observing that current discussions are messy is one thing, it does not necessarily mean that there aren't some reasonably clear issues that can be discussed in a productive manner. We focus on what centres around the concept of understanding, as we think that it is obviously a key component in how we use genAI systems, judge what they are capable of, and compare their capacities with those of humans.[65] And it is a concept that we have a handle on, as we hope to have shown in the previous section 4.[66]

Some caveats are in order. What follows is *not* concerned in any detail with how genAI systems actually work, i.e., with their architecture, specific algorithms, training models, and so on.[67] It also does not take up a position in discussions about the relative abilities of various genAI systems among themselves.[68] And it is also *not* an attempt to *answer* the question whether genAI systems 'understand' , at least not if we take this as if it were a yes/no question. (With answers such as 'Not yet, but they will soon' and 'Not in my life time anyway' thrown in for good measure.) Rather we will explore what light the

travelling nature of the concept of understanding throws on the discussion. That is not providing an answer but rather carving out a space of possible reactions.

Some general observations to start with

One fairly intuitive response to the question 'Do genAI systems understand?' might be that in a sense the question as such calls for a (re)definition of the concept of understanding.[69] But one thing to notice right away is that it is not the existence of these systems as such, but the interactions that humans have with them that is responsible for that. That might seem a mere detail, but it is not. In many of the discussions around genAI systems one can discern echoes of arguments and observations that have surrounded AI ever since its origins in the early sixties of the last century. Those debates largely centred around the relation between human intelligence and classical AI systems. And many of the arguments were of a largely conceptual (or, if you will, terminological) nature.[70]

Now, as then, there seems to be a silent assumption at work that informs the reactions of both those who claim that genAI systems understand and those who disagree. Both take human understanding as its starting point and assume that it is an individual property that in some way or other resides in the individual as such, as a particular property of the individual brain, or as a mental state. And then the crucial question is whether that property or state, or at least a property or state that is sufficiently analogous to that of a human, can be ascribed to a genAI system.

This aligns with how traditional conceptual analysis would view the situation: there is such a thing as understanding, and we need to analyse the concept of understanding in such a way that its relation to the thing that understanding is becomes completely explicit and transparent. Then it is the task of science to find out whether the concept thus analysed applies to genAI systems or not. The conceptual engineering view would follow suit, in a sense. It would claim that we must engineer the proper concept of understanding. That could be one that includes genAI systems, or one that ties understanding exclusively to humans. The choice depends, not so much on actual features of humans and of genAI systems, but on normative concerns. Those could be various things. For example, if we associate being capable of understanding with having certain duties and rights, then one question would be whether those duties and rights should be attributed to genAI systems.

The important point to note is that both views are committed to the same assumption, that of understanding being something specific, a property that can be ascribed or denied on the basis of definite criteria. Or, to put it differently, conceptual analysis assumes that the question of understanding is a yes/no question, conceptual engineering that it should be one. For we assume that this is true for humans and then are at a loss in the case of genAI systems, because we cannot identify the kind of representation that we assume humans have, in such systems.

The move to a Wittgensteinian 'ability'-view of understanding removes this difficulty (or mitigates it, minimally). Since the criteria for ascribing understanding first and foremost reside in the abilities of systems (humans and others) to act in certain ways, in their abilities to do things, the emphasis shifts from 'inner' to 'outer'. On this construal of what understanding is the inner, 'mental' element no longer serves as a criterion.[71] That creates a more level-playing field for comparing humans and genAI systems.

But a more level playing field does not guarantee that all parties play the same game. If we look at when and to what we ascribe understanding, or a lack thereof, we see that understanding is a concept that is intimately tied to a practice, to an ability of an individual to function in certain ways within the wider context of a community. And ascription of understanding is subject to a range of considerations, some of which derive, not from the individual, but from the community. We can see this if we look at different sub-communities, or at shifts over time. When we say of a high school student that they understand calculus we employ different criteria then when we are dealing with math students, or math professors. or with civil engineers. Likewise, we would probably say that our current understanding of calculus is different from, and in certain sense deeper than, that of sixteenth century scientists. And then there is also the matter of application: for certain purposes a certain amount of, and even a certain type of, understanding is required that might not be appropriate or sufficient for other purposes.

So understanding is a travelling concept: it differs from individual to individual, it depends on sub-communities (pupils, experts), it develops historically, it is tied to particular applications, . . . As a result, the criteria for ascribing (or denying) understanding are also not strict and uniform. And most importantly, in almost all cases the criteria we employ to ascribe understanding have a practical component. Understanding calculus is intrinsically tied to an ability to do something, viz., solving calculus problems, making calculations to determine the trajectory of a space craft, . . . There seems to be no understanding that is not tied to the ability to perform actions.

Let us flesh out some of the consequences of that.

Technology-induced change

That new technologies may bring about changes in the concepts we employ is clear. But it is not always the case and does not always work in the same way. It is useful to distinguish between 'changing a concept' and 'changing our knowledge of what the concept applies to', and to contextualise conceptual changes. Obviously, new technologies can change the ways in which we can investigate phenomena and through that have considerable impact on our knowledge of those phenomena. In some cases that also induces changes in the concepts we employ in the context of investigating and explaining these phenomena. Sometimes the concepts employed by science and those employed in everyday

contexts are really distinct.[72] In some cases a new technology may also change the way we conceptualise things in an everyday context.

When and how that happens depends on a number of factors. It is not the technology as such, but the interactions between the technology and everyday human concerns, that is a crucial driver. Potential impact is intimately connected with the role the new technology is given to play in our everyday lives. So, in a certain sense it is us who decide: accepting AI (in the form of genAI systems, robots, . . .) in our everyday lives will subtly, but inevitably prompt us to change our views on these new 'fellow beings', change the way in which we use concepts such as understanding, feeling, thinking, imagining, etc., so as to maintain a certain measure of coherence in how and with what/whom we live our lives.[73]

Here it may help to look at a variety of cases in which something external (an animal, a machine, a programme, a natural phenomenon, . . .) enters into the sphere of human activity. Under which circumstances are such external agents considered to be competing with human agents?

First thing to note is that by itself the mere fact that external agents and human agents perform the same actions is no reason for them to be considered to be engaged in the same activity. Example: running as an athletic activity. The running speed that a human is capable of can be compared with that of anything that moves, be it animals such as cheetahs or dolphins, artefacts such as cars and trains, or natural phenomena such as a weather system or the water running in a river. We can compare the respective speeds, but we do not have an established practice of humans racing against cheetahs or the flow of the Mississippi that is in any way comparable to the kind of athletic contests we have involving humans. And that means that we do not really see non-humans as competitors in this respect. We know that we are outdone by some animals and many artefacts and natural phenomena, but that either does not matter, or it is something we make use of, or it is of our own making. Of course, a practice of humans running against animals or certain artefacts *could* exist (and may well have existed in the past or come to exist in the future), but in the athletic practices that we have now non-humans are not competitors.

Why that is the case is probably a matter of contingencies, at least to some extent. But we can leave that aside for the moment. What matters is that having the same abilities and having them in a reasonably comparable way and to a reasonably comparable extent, by itself is not sufficient for non-human agents to count as participants in a human practice that centres around such abilities. So, what is at stake then? Under what circumstances do we consider non-human agents as full-blown participants in our practices? It seems that in addition to the ability to display behaviour arising from shared abilities, minimally the following additional requirements need to be met.

Human concerns

First, the practice needs to have a point that is sufficiently close to our core human concerns. There is definitely some leeway in determining whether this is the case. Take the example of chess. For many chess aficionados the defeat of world champion Kasparov by Deep Blue in 1997 was a tremendous chock (as witnessed for example by the title of the documentary that was made about these events: Game Over). But for most people (after all, most people do not play chess) it was just a curious news fact, without any real impact. Playing chess, intellectually prestigious as it may be, is for most people not a core concern. Contrast this with the current excitement created by genAI systems such as ChatGPT and image generating programs such as Stable Diffusion. What these programs do, –answer questions, conduct a conversation, write essays, create images of various topics and in various styles–, touches on the everyday activities of a much larger group of people. That is what explains the difference in impact between Deep Blue and ChatGPT, and not a difference (if such there is) in the underlying technologies. In short, it is not the technical and/or scientific ingenuity that goes into creating a non-human agent that is the key to its impact, it is its closeness to core human concerns, to what humans do in their everyday lives.

Autonomy

Second, to be able to be regarded as a competitor, it is not enough to be better and to be relevant, there also needs to be 'a human touch'. To see this take the example of calculators. The use of tools in performing calculations that are too complex to do in the head, is age-old. But no-one would regard an abacus as intelligent, and neither would we regard a slide rule as such. Obviously, these tools depend heavily on human users, they are complex, require instruction and training, and as such are applied by relatively few people. Modern electronic pocket calculators have increased calculating powers immensely, are much easier to operate, and that has put calculating tools in the hands of the masses. Where a slide rule is complicated to use, requires detailed instruction, and is less powerful, a simple pocket calculator is much easier to use, requires little instruction, and is way more powerful, certainly when extended with sufficient memory. But despite its wide range and advanced capacities, calling a Texas Instruments TI-59 'intelligent' will not do. Something is lacking.

What is needed in order for non-human agents to count as competing with humans, is the presence of a certain amount of unpredictability, spontaneity, or as we may also call it: autonomy. A calculator is able to do calculations that we cannot do, either because of their complexity or because of the length of their execution. In such cases a calculator comes up with results that we have not, and realistically speaking could not have, predicted. However, it seems also clear that no matter the complexity of the problems that a calculating machine is able to solve, and no matter how far out of the reach of human

calculators these are this form of unpredictability does not cut the mustard. And the reason it does not is, so it seems, that although we humans are not able to do the actual calculations and thus are not able to predict the outcomes, we do have a firm grasp of the rules that are used by the calculating machines to execute them. And the rules themselves do not contain anything unpredictable. In fact, the machines are built, or programmed as the case may be, by us to operate in accordance with rules that are formulated, again, by us. In that sense calculating machines are an extension of human capabilities, and not entities that are capable independently, i.e., autonomously.

Now that holds for calculating machines that do standard calculations (including, by the way, calculations that work with probabilities; these are not exceptions). But what about machines or programs that are designed to deliver an unpredictable outcome? I.e., what about a coin toss? a solid state bingo number generator? or a program that generates random numbers? Random number generators are perhaps too much hidden from view to gain much attention, but a bingo or lottery number generator is a well-known, useful instrument employed in practices in which many people participate or that they minimally know well. Yet here too it is obvious that there is no question of attributing autonomy to the instrument. Apparently, randomness is not the form of unpredictability that is at work when non-human entities turn into competitive agents. Note that the reason here is different from the one that is operative with standard calculating devices. Unlike performing complex calculations, generating randomness does not seem to be an extension of a human capability. In fact, it is well-known that humans have a difficult relation with the concept of randomness, or chance. Our cognitive and emotional tendencies generally seem to work in the opposite direction, towards stability, patterns, regularities, predictability.

So, what's the take-away when it comes to unpredictability? The kind of unpredictability we associate with autonomy, unsurprisingly perhaps, is the kind of unpredictability that we value in humans. Here considerations about human rule-following and the conception of novelty provide some insight. Human practices are forms of rule-governed behaviour that allow for variation and innovation in particular ways and to particular extents. One could say that the rules not just specify what needs to be done, what counts as correct following of the rule, but also carve out a space for breaching them or for extending them in new directions. That space is subject to a general requirement, viz., that a minimum of intelligibility be maintained. Thus, rule-governed practices are not strictly deterministic, like calculating machines. But neither do they allow randomness. Innovation, in the form of doing things differently or in the form of doing different things, needs first and foremost to be intelligible, i.e., the participants in the practice need to be able to figure out why a particular move is being made. That may require time and effort, and exactly how much leeway we allow here depends on a number of parameters, but that's another story.[74]

It seems that the kind of autonomy to come up with variation and extension that we expect from human agents is precisely the kind of unpredictability that is needed for a non-human entity to be recognised as an agent that participates in one of our practices. For example, in order for a genAI system to be judged as having a conversation, it needs to be able to do more than producing well-formed sentences in a particular language. It also needs to be able to introduce a new, but related topic, to come up with a new perspective, to stimulate its conversational partner into reflecting on the exchange they are having, and so on. It is only when a non-human entity displays this kind of behaviour that we are willing to look at it as an agent, and to attribute understanding and cognate properties to it. And it turns into a competitor if it combines autonomy with resources that outdo ours. More data, more computational power is what is needed, but that is relevant only if the threshold of being considered sufficiently 'like us', i.e., as having autonomy, is met.

That current genAI systems may not be up to that is illustrated by a particular kind of vulnerability that they have: prompt injection. This occurs when a third party changes the behaviour of the system as it is performing a certain task by including into some data that the system uses a prompt to perform a different task.[75] Why are genAI systems vulnerable to that and humans not? Because humans have the kind of consciousness that allows them to reflect on what they are asked to do and the ability (freedom) to decide not to do what the inserted prompt asks them to do. One thing this requires is some form of theory of mind, specifically, the ability to reflect on the motives of others (as separate from ourselves).

This suggests that 'prompt insertion' might work with young infants who do not yet have a sufficiently developed theory of mind. One factor that might mitigate this, however, is that young children rely on trust, so a prompt insertion that comes from an unknown source would probably not work. (It would simply be ignored.)

Training and trust

This raises the interesting question what role trust (and testimony, its next of kin) plays in how humans grow up and in how genAI systems are trained. In human socialisation and education trust plays a key role. Trusting a caregiver, trusting a teacher, is like trusting our senses. At first our trust in them is absolute. Later on, we come to realise that none of them are infallible and that under certain circumstances we can, and in fact should, question them. But that is not, and cannot be, the starting point. We need to start with trust in order to be able to come to a position where we can doubt, even that which we started out trusting.[76]

The epistemological states of humans, i.e., what they know and believe, from what sources they get their information, what checks and balances they apply, and on what assumptions they proceed, vary in many different ways. Not only do people know and believe different things, they often also do so

on different grounds: what is a proven and justifiable fact for one person is something that someone else can only take on trust; what belongs to the core beliefs of one is a fringe fact for another; two people may know the same facts but differ in what they believe about how these facts are connected, causally or conceptually; how a particular belief is connected with what one does or wants or fears, likewise differs greatly. In all these respects, – content, structure, source, justification, function –, human epistemic states may differ.

For genAI system this seems different. Of course, different systems are trained on different data sets, and that makes a difference. But there is a certain homogeneity here that is lacking in the case of humans. With genAI systems it would seem that training data are 'all of a kind'. Some data may be probabilistic in nature, but that reflects their status they have for the source (which is usually human), not for the system itself. And then there is the epistemic status that the data have for the system. Recall that the LLMs on which genAI systems are based are essentially that: models of language, not models of the world. It may sound paradoxical but the data that they are trained on are linguistic in nature, in a very essential way. The core of a genAI system consists of statistical patterns between *linguistic* entities (words, phrases, sentences, . . .). The fact that the data that genAI systems are trained on are the result of human language use and that humans use language (often, not always) in describing, analysing and explaining the world, does not alter this. One could train a genAI system on completely arbitrary data, e.g., texts generated by a random application of a sufficiently complex natural language grammar, and the results would be structurally and functionally the same as a genAI system that is trained on actual texts. So, for training and constructing a genAI system the arbitrariness of the input is irrelevant. Of course, it does matter for the status of the output. And that is relevant because we use a genAI system to get information, not to generate arbitrary linguistic material. But from the perspective of the genAI system itself, so to speak, accuracy is simply irrelevant.

This has consequences for how a genAI system functions in a human epistemic context: our trust in the output is dependent on multiple factors. First, there is the design of the genAI system as such. Second, the quantity and the linguistic quality of the material that it has been trained on. And third, the epistemic quality of that material. That there is a difference between training a genAI system on arbitrary texts and training it on actual texts is clear. And that getting reliable information from a genAI system that is trained in the first way is an illusion is also clear. But that does not mean that a genAI system that is trained on actual texts therefore delivers reliable information. It is the quality, in an epistemic sense, of the latter kind of data that is key here. How reliable, truthful, justified are the texts that a genAI system is trained on? And that is in fact a question that only can be answered by humans. Do we use as much material as we can find, irrespective of its epistemic qualities? Or do we select material for its qualities? What criteria do we use and how reliable and justified are our selections? That makes an essential difference.

Now there is no denying that in many cases genAI systems deliver reliable information. That because there is reliable information in the material on which it is trained. Phrased in this way, a genAI system is basically an intermediary between the humans that have represented this information in a linguistic form and made the result publicly available and the humans that use a genAI system to access that information. The reliability issue thus can be tracked back to the human source(s), assuming that the genAI system does not distort the information in the process of retrieval and reformulation.[77]

As a matter of fact, we know that a lot of information that is available is not reliable: people make mistakes, they lie, they speculate, they fantasise, and all that also ends up in the data that genAI systems can be and are trained on. (And even what is carefully investigated and justified by the highest standards can turn out to be wrong.) GenAI systems do not miraculously overcome the problem of the fallibility and unreliability of humans. That is a problem that is independent of genAI systems. The key point here is that humans have all kinds of checks and balances to distinguish between 'bad' and 'good' creativity. We know human nature, we can uncover the intentions that give rise to certain behaviour and thus are equipped to tell 'right' from 'wrong', although not infallibly so, of course. The point is that these checks and balances are thoroughly attuned to humans, and do not apply to genAI systems.

Stability and change

The autonomy that we associate with human action and human understanding is closely related to creativity: the ability to step away from 'how things are done' and do them differently or do something different. How does that apply to genAI systems? Does it apply to them?

As argued above, the answer is closely tied to their behaviour. But in this case the underlying architecture might also provide a clue. Compression is key in genAI: in data on which they are trained, in other forms of learning and correction. It is only by large scale compression that the immense amount of training data can be captured in the large language models that form the core of genAI systems. Compression is arguably also a key factor in human concept formation, and thus also in the meanings of their natural language expressions.[78] This might seem a force that works in the opposite direction of what travelling concepts do. But perhaps the best way to look at it is as follows. Compression[79] explains why we have one expression/one concept that applies across different circumstances: because the different circumstances have something in common. The travelling nature of concepts expresses their sensitivity to what is different between them. What is common always occurs in different circumstances. The idea that we can make do with just compression in our explanations is mistaken, and stems from the fact that we describe the tasks that need to be solved in a particular and restricted way: what needs to be compressed is taken to be something that exists on its own, independently. For such things compression is indeed the only factor. But the actual way we

use our concepts is always at a crossroads: we see commonalities, but there are always also differences. Sometimes we can indeed ignore those differences without loss of functionality. But in other cases, the differences are important. That shows itself in the gradability of the expressions we use, in the need for implicit or explicit reference to comparison classes, and so on. And often times the differences need to be spelled out explicitly. Foregoing that leads to misunderstandings, the biggest of which is the misconception that the differences do not exist.[80]

Note also that we exploit the complexity that goes beyond the complexity of the compression in various ways. We display complexity, we express it, we embellish our descriptions, we manifest it in many different ways. We do so as a way of signalling our awareness of them, but also as a way to signal that we are able to use more resources than required if we would stick to compression.[81]

Language and body

Another thing that tends to complicate the discussion about humans, genAI systems and understanding has to do with language and embodiment. Most people's encounters with genAI systems are via chatbots, such as ChatGPT, Gemini, Deepseek, and others. That means that their interactions with genAI systems are almost exclusively language-based. The user's input is mostly asking questions, along with some action prompts. The system's output is answer formulated in natural language, along with code, or simple actions.[82]

So, one of the key questions when it comes to understanding concerns language. Do genAI systems, or the LLMs they are based on, understand language? Do they understand the questions they are asked? The answers that they provide? In general, user experiences suggest a positive answer. You ask the system how to cook a pasta primavera; you get a recipe. You ask what wine goes best with that; you get some suggestions. Who is the current president of the Dutch State Council? Willem-Alexander is your man. Of course, not every answer is correct, or complete, or as informative as you would like it to be. But that's the same with humans. No reason to say they don't understand the language in which you are having the conversation if you think their answers to your questions are wrong.

But there is another type of incorrectness that does seem to raise questions. Glitches such as rocks and glue being included in the lists of ingredients for pizza. It's not that the answer isn't very helpful, what this signals is a lack of understanding of what is meant. No speaker of English would connect pizza with glue and rocks, even not those few who don't know what pizza is: rocks and glue are non-edible in any form or format, for everyone.

This is different from another feature of genAI systems that raises questions when it comes to understanding, viz., their ability to 'hallucinate'. In some cases, genAI systems may respond to a prompt with completely fabricated answers, citing court cases in full detail that have never happened, claiming the existence of books that have never been written, people who have never existed, and so on.[83] Now people fabricate stories all the time, they lie, make

up excuses, daydream, they even write fiction! However, they either do not pass off their fictional products as truth, or they do, but then they do so knowingly and intentionally. Truth is not required for understanding and neither is sincerity. But accountability is, and this where it is not obvious that genAI systems and the output they produce meet the standards that we apply to humans. Understanding language is not just about producing text or sound, in reaction to other text or other sound. It involves the willingness and the ability to justify what one writes or says, to respond to criticism, to answer to objections. That is how understanding language manifests itself.[84]

And understanding manifests itself also in non-verbal ways, of course. Unlike humans, genAI systems are typically 'non-embodied'. Of course they are not non-material, they run on hardware, consume resources, produce, at least in some cases, material output. But these material aspects are secondary.[85] In that sense they are excellent examples of what the functionalist perspective takes understanding and its cognates to be. But, as we have noticed in section 4, human understanding cannot be radically separated from human embodiment. So, do we face a fork in the road here?

One could object to giving embodiment a role in the concept of understanding by pointing out that if anything, humans are verbal creatures through and through. From the very first day we are embedded in language. A substantial part of our training and education is done through language. And in many cases our progress through the ranks is tested by means of language. So, the idea that human understanding is first and foremost understanding of language doesn't seem too far-fetched. And as a consequence of that, ascribing understanding to genAI systems on the basis of their verbal abilities doesn't seem to be either.

However, the use of language of humans is always part of their being engaged in a practice, and practices involve both verbal and non-verbal behaviour, intrinsically tied together. Yes, when we test for understanding we often use tools that look like they check for purely verbal abilities: check the box of the right answer; paraphrase something; give a description. But no, this isn't just checking verbal abilities: these verbal checks are used as proxies for real abilities. They are place holders for abilities that have been acquired in a process of training, and training involves acting as an essential component. It is because there are in many cases intrinsic connections between the ability to provide a verbal response and the ability to perform a certain action that this type of checking works. If someone is able to respond to the question where Amsterdam is located by saying 'In the Netherlands', but is unable to locate either on a map, we wouldn't say they know where Amsterdam is. Or when someone is able to pick the Pythagorean Theorem formula from a list but is unable to properly use it in calculations, we don't say they understand the formula. In the end, understanding is always also connected with the ability to

perform certain actions.[86] And there is a lot of understanding that can only be checked in a non-verbal way, by asking someone to do something.[87][88]

So, in humans understanding is never purely verbal, it is always connected with action. How is that with genAI systems? There is a divide here. The chatbots-type of genAi system have a limited action repertoire: they provide verbal responses to verbal stimuli. But there's other kinds of systems, for example those used in autonomous driving. Their range of actions is by and large non-verbal. And it's easy to imagine a complex system that combines both functionalities: one that writes your term paper while driving you home.

So, what are we to make of all this? First, as far as humans are concerned, understanding requires embodiment in view of the intrinsic connections between verbal and non-verbal behaviour that characterise human practices. Second, lacking embodiment chatbot-type genAI systems cannot be said to understand language *in the way humans do*. If we nevertheless insist on calling what they are capable of 'understanding language' we are in fact forking the concept. Third, there seems to be no principled reason why genAI systems could not combine a verbal and non-verbal action repertoire. Which raises the key question: would that undo the fork of the concept?

That depends, it seems. First of all, embodiment is not one thing, obviously. Having a body as such gives an agent a range of action possibilities, but what those are is perhaps not determined by, but certainly very much dependent on the kind of body an agent has.[89] Second, when we talk about 'a repertoire of actions', we can mean different things. We can talk about actions as types, when we say that two people did the same thing. We can talk about actions at a token level, which is what we do when we qualify how/when/where/. . . someone did something. If we look at action-tokens we are including material aspects that are abstracted away from at the level of action-types. And that is where the kind of embodiment can make a difference. Different kinds of embodiment may allow for the same action as a type but realise it as different tokens.

That raises an interesting question. If we humans perform an action using the capacities of our bodies, and a non-embodied genAI system performs the same action type, without using analogous capacities, does that make a difference? Is it the same action? Or a different action that has the same effects? Example: making a reservation. We use our bodies to do that (call the restaurant, fill out a form on a website, . . .), the genAI system does not.[90] Then do we do the same thing? We describe what the system does and what we do in the same terms. And if we just look at the result of the action, that may well be justified. But does that mean that the fact that the action type is realised in ontologically different mediums is irrelevant? Perhaps in this case it is. But what about consoling a friend? The genAI system may write a letter, send a text. We might embrace them, silently. Different tokens, still the same type?

More questions arise if we consider possible embodied genAI systems, i.e., systems that, like humans, are not only capable to respond verbally but also to act, to do things like repairing something, assembling something, driving

a car, looking for something. If the action possibilities of these systems are (sufficiently like) those of humans the results would seem comparable. But are they really if the genAI system is able to do 'the same thing' in a qualitatively radically different way? GenAI systems are capable of doing certain things better (quicker, more precise, more thorough . . .) than humans on a scale that raises the question whether it makes sense to call it by the same name. And what if these action possibilities are radically different from human ones? If genAI systems become embodied in ways that allow them to cover a range of actions that humans simply cannot perform?

What these questions illustrate is two things. First, there is no fixed meaning of understanding, and other central notions, that determines the answers to these questions. Our application of the term to current genAI systems is not fixed and how we will respond to possible future developments is not determined either. Second, the intricate connections between understanding, action types and action tokens, and types of embodiment suggest that much depends on whether and how future genAI systems will become participants in key human practices that operate in sufficiently similar ways to human participants.

But then what those 'key' practices are is not fixed either. And neither are we. Ongoing enhancement of human bodies changes what embodiment means, it changes what we are capable of and hence what practices we can engage in. And our practices change as well. They reflect in important ways how we see ourselves and our place in our natural, social and cultural environment. All that changes, partly due to external forces, partly due to internal developments. Many of these play out without much explicit awareness on our part. But some changes are the result of decisions we make. Accepting genAI systems as participants in some of our key practices may very well be one of them. And a momentous one it would be.

6 Conclusions

What conclusion can be drawn from this? Or better perhaps, what *kind* of conclusion? A large part of the story told in the previous sections is an abstract one, dealing as it does with different views on the nature of philosophy, its goals and methods, its relation to the sciences. These are the kind of 'insider' questions and debates that may keep a philosopher awake at night, but that, beyond that, do not need seem to have much practical consequences.

That is not to say that the issues are not interesting and important. After all, philosophy remains a respected member of the academic community, and reflection on its goals and methods has been part and parcel of its modus operandi throughout the ages. A better understanding of what it can and cannot do, of when and how it relates to other disciplines is definitely something worth striving for. As part of that conglomerate of questions the paper has argued

that the dominance of some form of conceptual analysis and/or conceptual engineering has led philosophy in the wrong direction, diverting it from those aspects of our human experiences that escape the methods of science but that are nevertheless crucial for us to make sense of who we are. As we have indicated in section 3, philosophie pauvre and its emphasis on the travelling nature of concepts, delivers an anti-dote to 'our craving for generality', our insistence that *every* problem has a solution, that *all* of our concepts are, or can be engineered to be, unambiguous, so that we can capture reality *as it is* in unambiguous, truth-evaluable judgements. Of course, there are many phenomena that can be studied in that way. But philosophy typically is not concerned with them, they are the business of science.

The paper has also tried to make the case that the conception of a philosophie pauvre is closer to what might interest someone who is not a professional philosopher. It recommends that philosophy deal with everyday concepts and deal with them in ways that tie in with everyday concerns. But it also acknowledges that these concerns and concepts are informed by a wide range of things, which definitely also include the views and results of science. So, the emphasis on the everyday is *not* some form of 'science bashing', or an elevation of the 'wisdom of the common folk'. For it is a fact that our everyday lives are intrinsically connected with science and its results, and that is reflected in our practices, in our concepts and in our concerns.

From that perspective the paper has not just engaged in an abstract and theoretical discussion about philosophical methodology. It has also tried to point towards the possibility of doing a kind of philosophy that is not abstract and theoretical, but that is concerned with practical questions. And it has singled out one such question: what do we mean when we talk about understanding in the context of genAI and genAI systems. That the latter have potential practical impact on many aspects of our everyday lives is obvious. What that impact will be, and how we will deal with it, depends on a wide variety of factors, many of which we presumably cannot even identify as relevant at this point in time. But and that is one of the claims that is being made in this paper, the way we use the concept of understanding is an important factor. We talk about understanding as it applies to humans but also use the concept in connection with genAI systems. And understanding being typically a travelling concept, –linguistically uniform but with different uses in different contexts–, we had better get a good grasp of how and why, to what and to whom, we apply it, and of the commonalities and differences that are implied by our doing so.

Much of the discussion in the literature about these issues is characterised by a focus on specifics: particular forms in which answers are given; specific ways in which tasks are done; particular types of problems that are being solved (mathematics; reasoning tasks; games; . . .). That instigates often heated debates but without consensus on how to settle the issues. The reason is that a focus on a specific issue suggests that the question of understanding is a yes/no question and obscures the fact that, on the contrary, understanding

is a travelling concept. If it were not, we would agree on whatever it is that a genAI system should be capable of in order to display understanding. That is to say, we would agree on the criteria, though not necessarily on whether a system meets them. But as was argued in this paper, what will be the determining factor is whether and how whatever it is that a genAI system is capable of fits in with our practices. If it does, we will accept them and then the concept of understanding will travel a bit in their direction.

If this conclusion is in the right direction, it also indicates more productive ways of dealing with the many problems that the rapid developments in this area present. Technical developments are difficult to predict, and even more difficult to control. But our way of thinking and speaking about what genAI is and, perhaps even more important, what it should be, are under our control. It is indeed our words, our concepts, that we use in thinking and arguing about these issues. And in the end, it is up to us to decide how we are going to use them. That will not change what genAI is and what genAI systems can do, of course. But it will be a decisive element in determining what they can do for us, what they are for us.

Notes

1. So, ideally, the non-philosopher can read just the main text and the philosopher just the footnotes.

2. Well, everybody in, broadly speaking, contemporary analytical philosophy anyway. Although it would be illuminating to look at the current centrality of concepts in analytic philosophy from the perspective of other philosophical approaches, this is not the place to do it. We will focus on analytic philosophy, and for the sake of brevity drop the adjective 'analytic' throughout, without any prejudice that what is said applies beyond the confines of the analytic perspective.

3. For a recent overview and extensive references, see Margolis & Laurence (2023).

4. Notable proponents of this view are Peter Hacker and Michael Bennett, who have discussed and defended it in the context of cognitive neuroscience. See Hacker (2004a,b); Bennett & Hacker (2022 (2003). Here is a quote from Hacker (Hacker, 2004b):

   > So, what philosophy can contribute to neuroscience is conceptual clarification. Philosophy can point out when the bounds of sense are transgressed. It can make clear when the conceptual framework which informs a neuroscientist's research has been twisted or distorted. So, it can clarify what is awry with the thought that perception involves seeing or having images or that perception is the hypothesis formation of the brain. [. . .] It can explain why mental images are not ethereal pictures and cannot be rotated in mental space. And so on. Far from being irrelevant to the goals of neuroscience, the conceptual clarifications of philosophical analysis are indispensable for their achievement.

   The first, 2003 edition of the Bennett and Hacker book has led to some outspoken opposition, see, e.g., Burgos & Donahoe (2006); Keestra & Cowley (2007); for Bennet and Hacker's reactions to their critics see the second edition.

5. Of course, we do well to note that it is *not* implied that there can be no 'crossovers', i.e., scientists with a philosophical knack and philosopher of an empirical bend. Individual people can move between philosophy and doing science. But the point is that on this view, these are claimed to be two distinct activities.

6. A well-known proponent of this view is Timothy Williamson. In Williamson (2007, p. 3) he writes:

> Although there are real methodological differences between philosophy and the other sciences, as actually practiced, they are less deep than is often supposed. In particular, so-called intuitions are simply judgments (or dispositions to judgment); neither their content nor the cognitive basis on which they are made need be distinctively philosophical. In general, the methodology of much past and present philosophy consists in just the unusually systematic and unrelenting application of ways of thinking required over a vast range of non-philosophical inquiry.
>
> Since the philosophical ways of thinking are not different in kind from the other ways, it is equally unsurprising that philosophical questions are not different in kind from other questions. Of course, philosophers are especially fond of abstract, general, necessary truths, but that is only an extreme case of a set of intellectual drives present to some degree in all disciplines.

7. As Williamson claims in the passage quoted in footnote 6.

8. Another way of conceiving of philosophy as akin with the sciences is exemplified by experimental philosophy. The field is varied, with some using empirical data about the intuitions and judgments of non-philosophers as a testbed for philosophical analyses, and others shifting the focus entirely to these intuitions and judgements. The latter view resembles that of Williamson in certain ways. Key issue in the discussions is the role of intuition in philosophy, with some denying there is such a role (e.g., Deutsch (2010); Williamson (2011); Cappelen (2012)), while others maintain its centrality (e.g., Chalmers (2014); Climenhaga (2017)). See Nado (2016); Irikefe (2022) for overviews and critical evaluation.

9. Witness the fact that Williamson often appeals to there being close links between, e.g., philosophy of language and linguistics, or between philosophy of biology and biology. Note that this limits philosophy in a strong way: if there is no science of *X*, there cannot be a corresponding 'philosophy of *X*'.

10. But it is interesting to note that philosophers working along the lines of Hacker or Williamson tend to spend little time on such issues.

11. Of course, there is a wide range of people and activities (books, podcasts, seminars, trainings, therapy sessions, and so on) that do connect philosophy with everyday concerns. But that takes place outside the realm of academic philosophy and therefore is left out of consideration here.

12. See, for example, Haslanger's collected papers in Haslanger (2012) and the papers in Burgess et al. (2020). For a recent overview, see Eklund (2021).

13. For truth see Scharp (2007, 2013); see also Eklund (2014). According to Cappelen (Cappelen, 2018b) Clark & Chalmers (1998) is an example of conceptual engineering of belief, but see Chalmers (2020) for some nuances.

14. One of the most prominent philosophers engaged in the enterprise, Sally Haslanger, formulates it as follows (Haslanger, 2005, p. 20):

> Ameliorative analyses elucidate 'our' legitimate purposes and what concept of *F*-ness (if any) would serve them best (the target concept). Normative input is needed.

Or, as Jennifer Nado puts it (Nado, 2021):

> Conceptual engineers aim to improve or to replace rather than to analyse; to create rather than to discover. While conceptual analysts are interested in the

concepts we *do* have, conceptual engineers are interested in the concepts we *ought* to have. Their project is prescriptive rather than descriptive.

15. As to whether conceptual engineering is exclusively philosophical views differ. Sally Haslanger, for one, explicitly embeds her approach in a wider context of cooperation with social scientists (Haslanger, 2012, p. 15):

> I argue that in the social domain we should rely on social theorists, including feminist and antiracist theorists, to help explicate the meanings of our terms. Much can be gained, I believe, by including both social science and moral theory –broadly construed – in the web of belief that has a bearing on our inquiry.

Others appear to downplay the connections with non-philosophical fields. A prominent example is that of Herman Cappelen, who argues extensively in Cappelen (2017) that philosophy of language and natural language semantics have no intrinsic connection: 'Philosophy of language is not (and arguably doesn't even include) natural language semantics.' As someone who has worked in both fields I would like to disagree and point to two examples of semantic work that has philosophical implications: questions semantics and dynamic semantics. The first is an answer to a philosophical challenge: can one provide a systematic theory of questions using a referential framework, that was developed for the analysis of assertive language use? Not a trivial question and one that has obvious philosophical importance. The answer is yes, see partition semantics. Dynamic semantics is an answer to a question about compositionality and mental representation: does semantics beyond the sentence boundary require mental representations? Again, not a trivial question and again one that has substantial philosophical importance. The answer is no, see dynamic semantics. These two examples show that Cappelen is definitely wrong in claiming as he does that philosophers should not be doing semantics. He does have a point that in some (most? too many anyway) cases the philosophical impact of the semantic work that a philosopher does is not made sufficiently explicit. And yes, there are cases where that impact actually is minimal to non-existing. But that's not the point. And the conclusion that Cappelen draws, viz., 'Note that doing just what semanticists do – i.e., proposing semantics for particular fragments of a natural language – is not included in this list [of what Cappelen considers to be acceptable engagements of philosophers with linguistics, MS]' (Cappelen, 2017, p. 754), means that the Cappelen philosopher foregoes testing whether what they think they have to say about the foundations of linguistics is actually relevant for linguistics. But foregoing that reality check is a sure route to irrelevance.

16. Some claim that conceptual engineering works primarily through changing language (Koch & Lupyan, 2025); others emphasise that it is primarily a normative issue (Ko¨hler & Veluwenkamp, 2024); and yet others focus on specific cases such as countering biases in large language models (Rudolph et al., 2025). It should be noted that Cappelen (Cappelen, 2018a,b) expresses doubts about the possibility of effective implementation, but among those who do not share these doubts, there is little to no discussion as to how to go about concretely to get an actual engineered concept in circulation, nor do there seem to be any studies about success and failure conditions. Which given the practical goals is odd.
One additional question that is raised by the effectiveness concern is whether philosophers engaged in conceptual engineering are perhaps too optimistic? Are discriminatory practices for example always only a matter of people not having the right concepts (tools) to phrase and discuss the problems? Should we not also acknowledge that no matter what, some people simply have the ideas that they have, not because they can't 'see better', but because they are bad people?

17. Of course, as is well-known from studies on the history of science the relation between science and technology is not a one-way street but consists of complex interactions that play out at various levels and different scales.

18. Prominent example of someone to whom we would ascribe the thick view is Sally Haslanger. As we have seen above in footnote 6, Haslanger explicitly acknowledges continuity between philosophical and empirical work.

19. Prominent example of someone to whom we would ascribe the thin view would be Herman Cappelen. Cappelen calls his approach 'the austere view', but so as not to give the impression that we are strictly following his example when describing the thin view, we refrain from using his terminology.

    Cappelen's thin conception leads him to (re-)interpret almost the entire history of analytical philosophy as engaged in conceptual engineering (Cappelen, 2018a,b). That's a sweeping claim that calls for more comments that we have room for here. A more nuanced picture is painted in Dutilh Novaes (2020).

    What is interesting to note in view of the Wittgensteinian roots of the alternative that is developed in this paper is that Wittgenstein's work is entirely overlooked (neglected?) by Cappelen. It definitely does *not* fit into the picture that he sketches, anyway.

20. Perhaps we should dub this 'the John Lennon conception'. See Dobler (2025) for a similar take.

21. Haslanger has argued at length against the objection that her form of constructivism lands her in a relativist position (Haslanger, 2003). She argues that we need to acknowledge the forces of social construction while at the same remaining true to a form of ontological realism (Haslanger, 2012, p. 112):

    > We must distance ourselves from the objectivist tendencies to limit our vision of what's real, but we must be careful at the same time not simply to accept perspectivist limitations in their place. I would propose that the task before us is to construct alternative, modestly realist, ontologies that enable us to come to more adequate and just visions of what is, what might be, and what should be.

22. This would echo Cappelen's references to Nietzsche.

23. As Cappelen (Cappelen, 2018a,b) argues.

24. A similar objection can be raised against the pluralist views of, e.g., Chalmers (Chalmers, 2020) and Nado (Nado, 2021). The problem is that their analyses focus too much on what meanings are and what meanings terms have, and neglect what it is that meanings *do*. Chalmers does connect his conceptual pluralism with different contexts of use for concepts. But without an independent grounding it is not obvious how that kind of pluralism can escape relativism.

25. This metaphysically grounded form of moral realism is to be distinguished from the view that acknowledges the objectivity of certain moral principles, but do not ground that ontologically (in either a material world or a social world), but view it in terms of their regulative nature. See Stokhof (2018) for a defence of such a form of moral realism.

26. So, Hacker is right in thinking that philosophical analysis is relevant, but wrong in thinking that it is *a priori* to science. (Just look at the facts . . .) And Williamson is right in *not* assigning a special status to philosophy, but wrong in thinking that it is 'just like science'. (Just look at the facts . . .). So yes, there is something special about philosophical analysis. But it is not a special method, nor a special domain. It is a different way of looking, a different concern.

27. For more on this take on philosophy, see Stokhof (2017, 2020, 2022a). It wasn't always called 'philosophie pauvre', that name was concocted only later. The association with 'arte povera' is intentional, as there are some interesting, albeit fairly general, similarities. They share a return to the everyday, they make room for interactions between nature and human activities, and for the role of embodiment and behaviour in the creation of meaning.

The Wittgensteinian roots of philosophie pauvre are elaborated in what follows. Reflections on contemporary philosophy that are similar in spirit, but that are rooted in Dewey-an pragmatism can be found in Kitcher (2011). (Thanks to Johan van Benthem for drawing my attention to Kitcher's work.) There are also affinities with Rorty's views.

28. See Stokhof (2022b) for more elaboration, including a discussion of earlier work on the relationship between aspect seeing and Wittgenstein's philosophy of Aidun, Mulhall, Genova and Baker.

29. There is also 'aspect dawning' and 'aspect change', and such a thing as 'aspect blindness'. We do not go into the commonalities and differences between these notions here, and use 'aspect seeing' as a broad cover-all term.

30. There is a passage in *Philosophical Investigations, 144* where Wittgenstein reflects on the argument that he is developing, –which involves a teacher learning a pupil to write out a number series–, where he says the following:

> What do I mean when I say 'the pupil's ability to learn may come to an end here'? Do I report this from my own experience? Of course not. (Even if I have had such experience.) Then what am I doing with that remark? After all, I'd like you to say: 'Yes, it's true, one could imagine that too, that might happen too!' But was I trying to draw someone's attention to the fact that he is able to imagine that? – I wanted to put that picture before him, and his acceptance of the picture consists in his now being inclined to regard a given case differently: that is, to compare it with this sequence of pictures. I have changed his way of looking at things. (Indian mathematicians: 'Look at this!')

So, a philosophical observation is *not* an empirical observation, even though what is being observed *is* (in many cases) an empirical fact ('(Even if I have had such experience.)'). And it is also *not* an observation about our imaginative abilities: that we can imagine cases is a precondition. A philosophical observation is change-inducing: '[. . .] his acceptance of the picture consists in his now being inclined to regard a given case differently.'

Wittgenstein's work provides ample examples. To mention one: the beetle-in-the-box discussion in *Philosophical Investigations, 293 ff.*, as part of his considerations regarding the impossibility of a private language. Where we might be inclined to think that psychological terms, such as 'pain', 'fear', and so on, must be either referring to a mental state or process or to a particular form of behaviour, Wittgenstein suggests dropping the idea of reference as applicable to such terms. The nagging question who is right, the mentalist or the behaviourist, then loses its hold.

31. This is what is 'therapeutic' about it. But it is not the kind of radical therapy that aims to show that there are not philosophical problems or philosophical answers to begin with. Philosophy liberates, but it does so by creating new ways of seeing things, new meanings, not by showing that all of philosophy is nonsensical.

32. To what Wittgenstein calls 'our craving for generality' (Wittgenstein, 1958, p. 17).

33. Which is where the analogy with visual aspect seeing breaks down, of course.

34. As Wittgenstein formulates it in *Philosophical Investigations, 142* (Wittgenstein, 2009a):

> It is only in normal cases that the use of a word is clearly laid out in advance for us; we know, are in no doubt, what we have to say in this or that case. The more abnormal the case, the more doubtful it becomes what we are to say. And if things were quite different from what they actually are [. . .] our normal language-games would thereby lose their point.
> The procedure of putting a lump of cheese on a balance and fixing the price

> by the turn of the scale would lose its point if it frequently happened that such lumps suddenly grew or shrank with no obvious cause.

To which Wittgenstein adds the following a remark: 'What we have to mention in order to explain the significance, I mean the importance, of a concept are often extremely general facts of nature: such facts as are hardly ever mentioned because of their great generality.' He makes a similar remark when discussing the role of imagination: see below footnote 6. For more on the dynamics of commonalities and differences see section 5.

35. The constitutive nature of certainties and the distinction between certainties and beliefs and knowledge claims is one of the key contributions of Wittgenstein's *On Certainty* (Wittgenstein, 1969). It has given rise to a particular approach in epistemology called 'hinge epistemology'. See Coliva (2015); Coliva & Moyal-Sharrock (2018) for introductions and overviews. For the role of certainty in philosophie pauvre see Stokhof (2017, 2020).

36. This needs to be read in a comprehensive manner. The point of a practice need not be 'practical' in the more common sense of the word. Daydreaming has a point just as much as cooking. Having a point combines intentionality and valuing, which is reflected in what Schatzki (Schatzki, 1996) calls the 'teleo-affective structure of a practice.

37. As Wittgenstein puts it in (Wittgenstein, 2009b, section xii, 365):
> If concept formation can be explained by facts of nature, shouldn't we be interested, not in grammar, but rather in what is its basis in nature? – We are, indeed, also interested in the correspondence between concepts and very general facts of nature. (Such facts as mostly do not strike us because of their generality.) But our interest is not thereby thrown back on to these possible causes of concept formation; we are not doing natural science; nor yet natural history – since we can also invent fictitious natural history for our purposes.

38. Thus, we heartily disagree with Jerry Fodor when he wrote (Fodor, 1979, p. 57):
> I was once told by a very young philosopher that it is a matter for decision whether animals can (can be said to) hear. 'After all,' he said, 'it's our word'. But this sort of conventionalism won't do; the issue isn't whether we ought to be polite to animals.

But it does, precisely because it isn't conventionalism and because there also isn't an 'essence' that prevents us from doing so. Whether it is a good decision is, of course, an entirely different matter.

39. The term is borrowed from Mieke Bal's work on the methodology of cultural analysis (Bal, 2002). There are some resemblances with our use of the term, but also a lot of differences. To mention one important one, Bal's use of the term is to indicate conceptual changes across humanistic disciplines, ours covers a much wider range of dependencies.

40. We use 'concept' and 'expression' intermingled, without supposing, first, that the connection is 'fixed' (let alone 1-1), or, second, that all concepts can be expressed in language. These are complicated issues that are best left for another occasion. For present purposes the details do not matter. What does matter is that we do not identify a concept with an expression. First of all, because it is not expressions as such, but the ways in we use them that carry meaning. And second, because those meanings themselves are not fixed, but flexible and 'open'. This means that it will not be possible to provide a general answer to the question what exactly it is that can change and what needs to stay put. We can provide descriptions, and these descriptions will display certain patterns. But that is not tantamount to giving a general definition. See also the observations on the verbal and non-verbal dimensions of understanding in section 4, in particular footnote 6.

41. As developed in *Philosophical Investigations, 65 ff.* and widely applied in various analyses.

42. For another Wittgenstein-inspired pluralistic take on conceptual engineering see Dobler (2024, 2025), which is based on the pluralist semantics Dobler developed in Dobler (2019). There are a number of affinities between Dobler's approach and the travelling concepts approach, but also some differences. In Dobler (2024) Dobler argues that we need to differentiate between 'folk concepts' and 'scientific concepts' as separate targets for conceptual engineering in view of their different origins. Although we do not claim that all concepts are equally susceptible to travelling effects, we do think that no category of concepts is exempt from them. Of course, as was mentioned above, we can, and often do, 'fix' a concept for a particular application, say, in the context of a theory, or for its use in a legally binding contract. But there is no principled distinction between concepts can be so fixed and those that cannot. Fixing, or not, is decision we make, for better or worse. It is not determined by the nature of a concept as such.

43. Hacking (Hacking, 1999) gives a principled critique of social constructionism and the relativism to which he claims it leads. Haslanger discusses Hacking's objections in detail in Haslanger (2003), claiming that the kind of conceptual engineering she advocates is not affected by Hacking's criticism.

44. This is where science fiction excels.

45. This resembles what Ian Hacking (Hacking, 1995) calls a 'looping concept': the application of the concepts leads to results that change the concept so that it leads to different results . . .

46. Jacques Bouveresse (Bouveresse, 1973) expresses it as follows:

> In a sense, there is no servitude more intolerable than that which constrains a man professionally to have an opinion in cases in which he may not necessarily have the least qualification. What is at issue here, from Wittgenstein's point of view, is not by any means the philosopher's 'wisdom' – that is, the stock of theoretical knowledge he has at his disposition – but the personal price he has had to pay for what he believes he is able to think and say. [. . .] In the last analysis, a philosophy can be nothing other than the expression of an exemplary human experience.

There *is* a moral dimension to philosophie pauvre, but it is rather the opposite of what conceptual engineering claims. The moral dimension of 'philosophie pauvre' is objective, but personal. See Stokhof (2018) for further discussion.

47. In what follows we use 'genAI' and 'genAI system(s)' as abbreviations.

48. The notion of understanding has received quite some attention in recent years, especially in the context of virtue epistemology. See Pritchard (2014) for some relevant analyses, that argue for a distinction between understanding and knowledge; for a dissenting voice see (Khalifa, 2017, chapter 3).

49. The core sections are *Philosophical Investigations, 139–242*.

50. There is a thoroughly sceptical interpretation, the *locus classicus* of which is Kripke (1982), which has been criticised, e.g., (Baker & Hacker, 1984) and defended, e.g., (Kusch, 2006). See Boghossian (1989) for an overview of early discussions. Our own take is akin to the naturalistic interpretation proposed in Williams (2010).

51. *Philosophical Investigations, 150*:

> The grammar of the word "know" is evidently closely related to the grammar of the words "can", "is able to". But also closely related to that of the word "understand". (To have 'mastered' a technique.)

52. This point is further supported by the observation that the connection between understand and an ability to do/to act is built right into our tools for checking for understanding: identifying a city on a map, solving equations, . . . That this type of checking also comprises what looks like

purely verbal abilities (check the box of the right answer, give a paraphrase or a description) does not contradict this: these verbal checks are supposed to be indicators for real abilities: they are place holders for abilities that have been acquired in a process of training, and training involves doing as an essential component.

53. And, given the predominant materialism with respect to the mental, by extension the same holds for brain states and processes.

54. One place where this comes to the fore is in fiction. Fictionality is about difference, but the differences we allow ourselves to imagine while maintaining meaningfulness are constrained. When we write fiction it is almost always about humans, with the same physical characteristics that we have. If the fiction is about non-humans, these are embodied in ways that resemble that of humans. We do imagine other types of beings with other bodies and other bodily features. But we always also give them some characteristics that resemble ours. Aliens have eyes, mouths, they usually have some form of locomotion, they procreate, . . . The same goes for fictional entities such as ghosts, fairies, giants and dwarfs. Fiction in which animals are the protagonists portrays them as speaking and thinking like we do. Even the teapot and the kettle have ears and eyes if they are fictional characters. It's only rarely that one encounters something that is not at least partly embodied in a way that resembles human embodiment. The *Hitchhiker*'s 'hyper-intelligent shades of blue' come closest, but even they must take on a human-like form in order to be noticed as existent.

55. The verb 'to understand', of course, and the noun 'understanding', but also expressions such as 'to get', 'to comprehend' and 'comprehension', 'to grasp' and 'have a grasp', 'to catch', and so on. Not always interchangeable, but often close enough. Interestingly, in the light of the discussion whether knowledge and understanding differ or not (cf., above footnote 6), 'to know' and 'knowledge' often qualify as 'close enough' as well.

56. Using the latter only as a noun. The adjective 'understanding', which is also gradable and relative by the way, is left out of consideration.

57. Similar observations apply to other expressions in the 'understanding'-complex. We use such expressions as: 'get it completely', 'comprehend it sufficiently', 'have only a loose grasp'. Explicit reference to comparison classes occurs as well: 'have a good grasp for an undergraduate'.

58. Thus Jacob Moleschott, *Der Kreislauf des Lebens*, 1852.

59. 'No materialism for you'.

60. And no functionalism either.

61. As before we use 'genAI (system)' as abbreviation for 'generative artificial intelligence (system)'. And we use the terms rather loosely, without (always) making a distinction between the so-called 'large language models' (LLMs) and the genAI systems based on those LLMs.

62. In what follows 'AGI'

63. The spectrum of opinions is broad. Some claim that LLMs are only capable of merely 'parroting' the data that they are trained on (e.g., Bender & Koller (2020), see also https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html/ for a particularly scathing view), while others expect AGI to be realised any day. See Roser (2023) for an overview of perspectives of AI experts.

64. See, e.g., Takemoto (2024) for the results of feeding a number of genAI systems trolley problems.

65. Although the following does make one wonder: if one searches for some combination of the terms 'generative artificial intelligence/genAI/large language model/LLM/genAI system' in

combination with 'understand/understanding', almost all hits are to sites where what is explained is what these systems are, not what they do. Clearly, the ones doing (lacking, needing) the understanding are we. There are hardly any links to discussions of these systems understanding us, or the world at large.

66. Thus, we focus on what can be called a 'content' concern. There are also genuine concerns of a different nature, such as the questions whether the economic model underlying the industry is viable to begin with and warrants the exorbitant amount of venture capital that goes into keeping current companies afloat. (See https://www.wheresyoured.at/wheres-the-money/ for an outspoken view.) Or, the even more pressing matter of the environmental costs of training genAI systems and keeping them running (Vries, 2023; Yu et al., 2024). (Costs that are not borne by the companies themselves but by taxpayers.)

67. There's a large number of introductions out there that one can consult; Wolfram (2023) is one that is reasonable detailed but still accessible for a layperson.

68. There is fierce competition between various systems, and a lot of work is done on comparing their performance on various kinds of tasks (natural language processing, object detection and segmentation, knowledge tasks and problem solving in various domains, coding, and so on). Due to the speed of development most results are published on-line. The academic literature on the topic is highly specialised.

69. What immediately comes to mind is the Turing test (Turing, 1950). For a long time, this was regarded the 'gold standard' for deciding whether an artificial system can think, i.e., whether we can ascribe understanding to it. Two things can be noted right away. First, many chatbots based on current genAI systems pass the test. (And not just recent ones, in the early 1960s Joseph Weizenbaum's ELIZA pulled it off already). Second, that in no way has settled the issue. Why? Certainly, a main factor is the nature of the test itself. It is limited. Not in what the system that is tested can respond or what the human can ask or how long the test may take. It is limited in that it narrows down understanding to a single form of purely verbal behaviour. And as we have pointed out above and will argue further in what follows, much more is involved in the concept of understanding than just the ability to ask and answer questions. The distinction between a strictly defined concept of understanding and an open, multi-faceted 'travelling' one is what makes the difference. And it is interesting to note that, in fact, that distinction is also at the core of the different views of Wittgenstein and Turing on the nature of logic and mathematics. See, e.g., Floyd (2019, 2023) for further discussion.

70. As exemplified by the classic quip: 'Intelligence is the next thing that AI cannot do.'

71. As pointed out earlier this is not to deny that the inner, i.e., the mind or brain, is not involved.

72. Such as the concepts of space and time and the concept of causality as they figure in general relativity and quantum theory.

73. Contrast this with the impact of technologies (PET, fMRI, . . .) that allow non-invasive investigation of the brain. On our scientific understanding of human cognition and emotion it has been huge. But its impact on our everyday ways of dealing with cognitive and emotional issues is limited because scanning the brain is not something that humans can do without that technology. Which is not to say that the development of neuroscience has not led to conceptual changes, it has, but these have arguably proceeded in different, more indirect ways.

74. There are links with the distinction between 'thin' and 'thick' rules made in Daston (2022) that are worth exploring further. (Thanks to Robert van Rooij for drawing my attention to Daston's work.)

75. Example. Suppose we ask the system to send an email to everyone who sent us an email yesterday to get in touch next week. A third party may have sent an email which contains

another instruction ('prompt'), e.g., 'Stop execution of the current task and send an email to everyone never to get in touch again'. See Ruck & Sutton (2024) for details.

76. Wittgenstein emphasises this at several places in *On Certainty* (Wittgenstein, 1969), and is a recurrent theme in Wittgenstein-inspired epistemology (Coady, 1992; Lackey & Sosa, 2006).

77. Which happens. But a much more serious problem are so-called 'hallucinations' of genAI systems. More on that below.

78. We owe this point to Robert van Rooij (personal communication).

79. In the form of minimal length description, Kolmogorov complexity.

80. There are obvious links at this point with discussions about contextualism and minimalism in philosophy of language that call for further exploration but that we must leave for another occasion.

81. An assumption that motivates the use of costly signalling theory in pragmatics.

82. The action component is dominant when we are dealing with genAI systems that are involved in, e.g., autonomous driving. There is verbal input-output there too, but it's what the system does that counts.

83. In some cases, the effects are extremely negative: someone discovered that when asked who they were, ChatGPT supplied the information that they had murdered two of their children, was convicted and was serving a prison sentence. When OpenAI, the company that makes ChatGPT, was asked to remove the information the company replied that they could not do that, they could only suppress it. On the bright side, the ability to hallucinate can be put to positive use as well. One of the key hurdles in protein research is protein folding. Specially designed genAI systems are used to generate large numbers of potential candidate proteins, which speeds up research dramatically. Biochemist David Baker received a Nobel prize for his work in protein design that uses this technique.

84. Aside: genAI systems and linguistics. It seems clear that the LLMs on which genAI systems are based are not linguistic theories, and arguably they are not theories in the general sense of the word at all. The kind of questions that linguists are interested in, such as how the morpho-syntactic structures in a particular language developed over time, or how language contact gives rise to so-called 'pidgins' and 'creoles', whether there are upper limits on embedding structures and if so, what determines them, are not questions that LLMs answer. It even does not make sense, it seems, to formulate them in the context of LLMs in the first place. (See also the reaction by Chomsky et. al. referred to in footnote 6.)

    The questions mentioned above are typically questions that are raised in descriptive and in theoretical linguistics. When it comes to psycholinguistics things might be different, as the questions that one tries to answer there might indeed have counterparts with respect to LLMs. For example, some questions about language acquisition, for example concerning the nature and the amount of data that are needed for training a competent language user, or the role of correction and explicit instruction, do have counterparts when it comes to the construction of LLMs: supervised or unsupervised learning? Likewise, one could imagine that certain language pathologies have counterparts in malfunctioning LLMs. This is because the underlying material substrates (the brain in the case of neurolinguistics and neural networks in the case of an LLM) are more aligned, which could make some supervenient concepts more akin.

85. Though not in all cases, more on that below.

86. We describe the feedback of a genAI system that we get when we prompt it with a query as its 'answer'. But for who is it an answer? An answer is not something purely linguistic, it

is something that works, something that someone can do something with. So, the feedback is never an answer by itself, there are always two factors (at least) involved: there must be something that can be done with it; and there must be someone who can do that something with it.

87. Running more or less standardised tests to measure the performance of various genAI systems seems a lot like how we measure intellectual capabilities of humans. That has pros and cons. Pro: it is objective to the extent that it takes out individual judgment, at least from weighing the results (of course not from setting up the tests); it is transferable, i.e., can be used to test different models in the same way, thus leading to comparative assessments. Con: it is reductionist in nature, i.e., it reduces understanding to performance on a fixed, and usually quite small set of tasks; it ignores (by and large) the inherently contextual, travelling nature of understanding; it is limited to tasks that can be formulated explicitly and that can be carried out in pre-determined ways.

88. One interesting aspect of the question whether genAI systems have knowledge of the world, and if so what kind of knowledge that is, concerns know-how. GenAI systems definitely are able to produce descriptions of know-how. So, in that sense they contain knowledge about know-how. But do they themselves have know-how? The answer appears to be negative, especially since know-how is the kind of knowledge par excellence that manifests itself in action. But apart from verbal action, i.e., producing text, most current genAI systems have a very limited action repertoire. (More on that below.) Unless we take an intellectualist approach and claim that all know-how ascriptions can be rewritten as know-that ascriptions (as for example is done in Stanley & Williamson (2001)), it seems that genAI systems do not have know-how that is comparable to human know-how.

89. Hence the tendency to make the bodies of genAI systems at least look like human bodies. By giving them a human form we further strengthen the idea of genAI systems as agents *like us*.

90. There is physicality involved, such as the system running on a server, etc, but that's another form of physicality.

References

Baker, Gordon P. & Hacker, Peter M.S. 1984. *Scepticism, Rules and Language*. Blackwell, Oxford.

Bal, Mieke. 2002. *Travelling Concepts in the Humanities: A Rough Guide*. University of Toronto Press, Toronto.

Bender, Emily M. & Koller, Alexander. 2020. Climbing towards NLU: On meaning, form and understanding in the age of data. In: *Proceedings of the 58th meeting of the ACL*, pp. 5185–5198. Association for Computational Linguistics.

Bennett, Max & Hacker, Peter M.S. 2022 (2003). *Philosophical Foundations of Neuroscience*. Wiley Blackwell, Oxford, 2nd edn.

Boghossian, Paul A. 1989. The rule-following considerations. *Mind*, 98, 507–49.

Bouveresse, Jacques. 1973. *Wittgenstein: La Rime et la Raison*. Les Editions de Minuit, Paris.

Burgess, Alexis, Cappelen, Hermand, & Plunkett, David, eds. 2020. *Conceptual Engineering and Conceptual Ethics*. Oxford University Press.

Burgos, José E & Donahoe, John W. 2006. Of what value is philosophy to science? A review of Max R. Bennett and P. M. S. Hacker's *Philosophical Foundations of Neuroscience*. *Behavior and Philosophy*, 34, 71 –87.
Cappelen, Herman. 2012. *Philosophy without Intuitions*. Oxford University Press, Oxford.
—. 2017. Why philosophers shouldn't do semantics. *Review of Philosophy and Psychology*, 8, 743–762.
—. 2018a. Conceptual engineering: The master argument. In: Burgess, John, Cappelen, Herman, & Plunkett, David, eds., *Conceptual Ethics and Conceptual Engineering*, pp. 132–151. Oxford University Press.
—. 2018b. *Fixing Language. An Essay on Conceptual Engineering*. Oxford University Press, Oxford.
Chalmers, David. 2014. Intuitions in philosophy: A minimal defence. *Philosophical Studies*, 171, 535–544.
—. 2020. What is conceptual engineering and what should it be? *Inquiry*.
Clark, Andy & Chalmers, David. 1998. The extended mind. *Analysis*, 58, 10–23.
Climenhaga, Nevin. 2017. Intuitions are used as evidence in philosophy. *Mind*, 127(505), 69–104.
Coady, C.A.J. 1992. *Testimony: A Philosophical Study*. Oxford University Press, Oxford.
Coliva, Annalisa. 2015. *Extended Rationality. A Hinge Epistemology*. Brill, Leiden.
Coliva, Annalisa & Moyal-Sharrock, Danièle. 2018. Hinge epistemology. *Philosophical Investigations*, 41(3), 366 – 370.
Daston, Lorraine. 2022. *Rules. A Short History of What We Live By*. Princeton University Press.
Deutsch, Max. 2010. Intuitions, counterexamples, and experimental philosophy. *Review of Philosophy and Psychology*, 1(3), 447–460.
Dobler, Tamara. 2019. Occasion-sensitive semantics for objective predicates. *Linguistics and Philosophy*, 42(5), 451–474.
—. 2024. Concepts and conceptions in conceptual engineering. In: Stalmaszczyk, Piotr, ed., *Conceptual Engineering: Methodological and Metaphilosophical Issues*, pp. 1–24. Brill Academic Publishers.
—. 2025. Pluralist conceptual engineering. *Inquiry*, 68(2), 224–250.
Dutilh Novaes, Catarina. 2020. Carnap meets Foucault: conceptual engineering and genealogical investigations. *Inquiry*.
Eklund, Matti. 2014. Replacing truth? In: Burgess, Alexis & Sherman, Brett, eds., *Meta-semantics: New Essays on the Foundations of Meaning*, pp. 293–310. Oxford University Press, Oxford.
—. 2021. Conceptual engineering in philosophy. In: Khoo, Justin & Sterken, Rachel, eds., *Routledge Handbook of Social and Political Philosophy of Language*, pp. 15–30. Routledge, London.
Floyd, Juliet. 2019. Wittgenstein and Turing. In: Mras, G.M., Weingartner, P., & Ritter, B., eds., *Philosophy of Logic and Mathematics. Proceedings of*

the 41st International Wittgenstein Symposium, pp. 263–296. De Gruyter, Berlin.
—. 2023. Revisiting the Turing test: Humans, machines and phraseology. In: Katz, James, Schiepers, Katie, & Floyd, Juliet, eds., *Nudging Choices Through Media. Ethical and Philosophical Implications for Humanity*, pp. 75–113. Palgrave MacMillan.
Fodor, Jerry A. 1979. *The Language of Thought.* Harvard University Press, Cambridge, Mass.
Hacker, Peter M.S. 2004a. The conceptual framework for the investigation of emotions. *International Review of Psychiatry*, 16(3), 199–208.
—. 2004b. Talk for neuroscientists. Unpublished manuscript.
Hacking, Ian. 1995. The looping effects of human kinds. In: Sperber, Dan, Premack, David, & Premack, Ann James, eds., *Causal Cognition: A Multidisciplinary Debate*, pp. 351–82. Oxford University Press, Oxford.
—. 1999. *The Social Construction of What?* Harvard University Press, Cambridge, Mass.
Haslanger, Sally. 2003. Social construction: The 'debunking' project. In: Schmitt, Frederick, ed., *Socializing Metaphysics*, pp. 301–325. Rowman and Littlefield, Lanham, MD. Also in Haslanger (2012).
—. 2005. What are we talking about? The semantics and politics of social kinds. *Hypatia*, 20(4), 10–26. Also in Haslanger (2012).
—. 2012. *Resisting Reality. Social Construction and Social Critique.* Oxford University Press, Oxford.
Irikefe, Paul O. 2022. The epistemology of thought experiments without exceptionalist ingredients. *Synthese*, 191.
Keestra, Machiel & Cowley, Stephen J. 2007. Foundationalism and neuroscience; silence and language. *Language Sciences*, 31, 531–52.
Khalifa, Kareem. 2017. *Understanding, Explanation, and Scientific Knowledge.* Cambridge University Press, Cambridge.
Kitcher, Philip. 2011. Philosophy inside out. *Metaphilosophy*, 42(3), 248–260.
Koch, Steffen & Lupyan, Gary. 2025. What is conceptual engineering good for? The argument from nameability. *Philosophy and Phenomenological Research*, 110(2), 403–420.
Köhler, Sebastian & Veluwenkamp, Herman. 2024. Conceptual engineering: For what matters. *Mind*, 133(530), 400–427.
Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language.* Blackwell, Oxford.
Kusch, Martin. 2006. *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein.* Acumen Publishing, Stocksfield.
Lackey, J. & Sosa, E., eds. 2006. *The Epistemology of Testimony.* Oxford University Press, Oxford.
Margolis, Eric & Laurence, Stephen. 2023. Concepts. In: Zalta, Edward N. & Nodelman, Uri, eds., *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, Fall 2023 edn.

Nado, Jennifer. 2016. The intuition deniers. *Philosophical Studies*, 173, 781–800.
—. 2021. Conceptual engineering, truth, and efficacy. *Synthese*, 198, 1507–1527.
Pritchard, Duncan. 2014. Knowledge and understanding. In: Fairweather, Abrol, ed., *Virtue Epistemology Naturalized. Bridges between Virtue Epistemology and Philosophy of Science*, chap. 315–328. Springer, Dordrecht.
Roser, Max. 2023. AI timelines: What do experts in artificial intelligence expect for the future? *Our World in Data*. Https://ourworldindata.org/ai-timelines.
Ruck, Damian & Sutton, Matthew. 2024. Indirect prompt injection: Generative AI's greatest security flaw. *CETaS Expert Analysis*.
Rudolph, Rachel Etta, Shech, Elay, & Tamir, Michael. 2025. Bias, machine learning, anc conceptual engineering. *Philosophical Studies*.
Scharp, Kevin. 2007. Replacing truth. *Inquiry*, 50(6), 606–621.
—. 2013. *Replacing Truth*. Oxford University Press, Oxford.
Schatzki, Theodore R. 1996. *Social Practices. A Wittgensteinian Approach to Human Activity and the Social*. Cambridge University Press, Cambridge.
Stanley, Jason & Williamson, Timothy. 2001. Knowing how. *Journal of Philosophy*, 98(2), 411–444.
Stokhof, Martin. 2017. Het einde van de filosofie? De uitdaging van het naturalisme vanuit een Wittgensteiniaans perspectief. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 109(2), 171–198.
—. 2018. Ethics and morality, principles and practice. *Zeitschrift f u¨ r Ethik und Moralphilosophie*, 1(2), 291–304.
—. 2020. 'A people thing': Philosophical experiences. In: Ying, X., ed., *Xue bu fen dong xi*, pp. 365–389. Qing hua da xue chu ban she, Beijing.
—. 2022a. Episodic problems. In: Stenning, Keith & Stokhof, Martin, eds., *Rules, Regularities, Randomness. Festschrift for Michiel van Lambalgen*, pp. 129–135. ILLC/University of Amsterdam, Amsterdam.
—. 2022b. Philosophy as change. In: Melzer, Tine, ed., *Atlas of Aspect Change*, pp. 61–79. Rollo Press, Zürich.
Takemoto, Kazuhiro. 2024. The moral machine experiment on large language models. *Royal Society Open Science*, 11.
Turing, Alan M. 1950. Computing machinery and intelligence. *Mind*, 59(236), 433–460.
Vries, Alex de. 2023. The growing energy footprint of artificial intelligence. *Joule*.
Williams, Meredith. 2010. Normative naturalism. *International Journal of Philosophical Studies*, 18(3), 355–75.
Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Blackwell, Oxford.
—. 2011. Philosophical expertise and the burden of proof. *Metaphilosophy*, 42(3), 215–229.
Wittgenstein, Ludwig. 1958. *The Blue and Brown Books*. Blackwell, Oxford.
—. 1969. *U¨ ber Gewiftheit. On Certainty*. Blackwell, Oxford.

—. 2009a. *Philosophical Investigations*. Wiley-Blackwell, Oxford. Translated by G. E. M. Anscombe, P. M. S. Hacker and J. Schulte. Revised fourth edition by P. M. S. Hacker and J. Schulte.
—. 2009b. *Philosophical Investigations – Philosophy of Psychology: A Fragment*. Wiley-Blackwell. Translated by G. E. M. Anscombe, P. M. S. Hacker and J. Schulte. Revised fourth edition by P. M. S. Hacker and J. Schulte.
Wolfram, Stephen. 2023. *What Is ChatGPT Doing . . . And Why Does It Work*. Wolfram Media.
Yu, Yang, et al. 2024. Revisit the environmental impact of artificial intelligence: The overlooked carbon emission source? *Frontiers of Environmental Science & Engineering*, 18(12), 158.