How to be a good philosopher?
Are there any rules?*

Abstract: *The paper represents an attempt to present and discuss a puzzle derived from Ludwig Wittgenstein's rule-following considerations, aiming to avoid the existent ambiguities and conceptual misunderstandings. It is shown that the traditional logical approach to language is in principle unable to solve the puzzle. Although this result was taken to balance the odds in the favour of a naturalistic approach to the aim of producing a philosophical theory of natural language, it is also shown that this approach is equally wrong-headed. A few more general considerations with respect to the implications of this negative result are formulated in the end.*

1. It has been said more than once about philosophical questions that they are meaningless. And it might be true of at least some questions usually asked by philosophers that they are so indeed. As philosophers, we cannot just rely on our own natural linguistic abilities and hope that this will do. Several examples have been produced, if only in the philosophical literature of the last century, illustrating situations in which sentences that we would be inclined to consider meaningful prove to be, after a more careful examination, quite meaningless[1]. As a consequence, having a solid theory of language to provide us with the criteria for distinguishing between meaningful and meaningless sentences could be regarded as a prerequisite task for a philosophical investigation. I believe that such a theory will rest at least on the following assumption[2]:

> (A) There is a set of rules that govern our linguistic activities, such that, for any linguistic utterance, we are able to say if it breaks or agrees with some of these rules. We can point out which rules were supposed to apply to the considered situation and which, if this is the case, have not been correctly followed. In the second case, not only can we state that a rule that was supposed to be followed has been broken, but we can also reasonably argue, by making appeal to the rule, that it so happened.

Accordingly, the theory will aim to state the rules one must obey if one wishes to speak meaningfully. It will also provide a procedure to apply the rules to

* This paper was supported by a research grant offered by the Alexander von Humboldt Foundation between May and September 2002 in the frame of a special program for scientific recovery in Balkans. For useful comments and suggestions I am indebted to Mircea Flonta, Herbert Schnädelbach, Geert Keil and Kathrin Gluer-Pagin.

[1] Adrian Paul Iliescu (in "Wittgenstein and the problem of nonsensical philosophical questions", Rev. filos., XLV, 4, p. 459-470, Bucharest, 1998 [in Romanian]) discusses a few examples of this sort offered by Wittgenstein.

[2] I have elsewhere (namely, in Gheorghe Stefanov, The Wittgensteinian Challenge to the Contemporary Philosophical Approaches to Language, doctoral thesis submitted at the Department of Philosophy, University of Bucharest, 2000, [Romanian, unpublished]) considered other assumptions (like the one that we can distinguish between simple and complex elements of language, between the sense and the force of an utterance, between language and metalanguage or between logical rules and semantical rules) and tried to show how they could be criticised from a Wittgensteinean point of view. That enterprise was somewhat similar (keeping the proportions) to the one pursued by Hacker and Baker, 1984. However, I believe that (A) goes deeper than any other assumption. Indeed, it is hard to see how *any* normative semantical theory not assuming a version of (A) would look like.

each possible utterance, in order to see if it agrees with the rules or not. Were we armed with the set of rules and the procedure to apply them correctly to each case, we would be able to detect, for any question we might want to raise, whether it is meaningful or not. Thus, we could start a philosophical investigation being completely assured that the question we are trying to answer is not a meaningless one.

The problem is that (A) is wrong and it should be rejected. The most radical criticism of (A) is fostered by the later Wittgenstein's rule-following considerations. In what follows I will do little more (if any) than re-enact Wittgenstein's arguments. With respect to this, I do not wish to get into disputes regarding what Wittgenstein's rule-following considerations (RFC) really mean[3]. I would rather take it for granted that I understood Wittgenstein correctly. I am, however, mostly willing to accept that I did not if the arguments presented here turn out to be feeble. But if someone accepts them, then I think the whole credit should be given to Wittgenstein[4]. The structure of this paper is as follows. I will first show how Wittgenstein's RFC lead to the conclusion that if the rules are seen as some sort of entities (linguistic, abstract or perhaps mental), then it is not possible to bridge the gap between them and our linguistic practice. This, I think, is enough to show that the traditional logical/analytical approach to language is in principle wrong-headed. I will not insist on this point because it seems to me that it has been lately understood, and the consequences of Wittgenstein's arguments in this respect have been gradually gaining acceptance. The last part of my paper will be dedicated to showing that an attempt to offer a naturalistic explanation of the relation between semantic rules and linguistic activities, usually presented as an alternative to the analytical approach, is equally mistaken. A few general conclusions will be presented in the end.

2. Let us begin by looking at an example. Suppose that we have two extremely simple languages[5], L_1 and L_2. Each of them has only three sentences. For the purpose of this presentation there is no point in getting into the internal structure of these sentences so, since both L_1 and L_2 are finite, we could skip formation rules and define our languages in a descriptive manner:

L_1 - {'A', 'B', 'C'}
L_2 - {'1', '2', '3'}

[3] There are at least four different readings of Wittgenstein's RFC. One reading has started with Kripke, 1982. A different interpretation was initiated by Wright, 1981, pp. 99-117. A more recent one is due to McDowell, 1992. (see also McDowell, 1998, essay 11 - "Wittgenstein on Following a Rule".) Each of these has generated various attempts to provide a solution. For a systematic summary of the debate generated by Kripke's version see, for instance, Boghossian, 1989. Crispin Wright's version of the argument has received (*inter alia*) different replies from both McDowell and Gareth Evans (Evans, 1985). The "standard reading" (i.e. the one that I hope not to distance very much from) is the one offered by Hacker and Baker, 1985.

[4] Yet Wittgenstein or Wittgensteineans would not agree perhaps with some of the things I am saying here. This is due to the fact that my strategy is a more tolerant one (I am inclined to accept as much as I can of the criticised position, if I can still show that it is wrong).

[5] If one is reluctant to accept them as fully developed languages, then I could say that L_1 and L_2, together with the procedure of translating the sentences from L_1 to sentences from L_2 introduced below, form a particular language-game.

A text written in L_1 will look like this: ‘A. C. A. B. C.’[6]
The same text could be written in L_2 like this: ‘1. 3. 1. 2. 3.’

Now we state the following general rule for translating[7] from L_1 into L_2:

(R) The $\mathbf{n}^{th}$ sentence from L_1 is correctly translated into L_2 if it is replaced by the $\mathbf{n}^{th}$ sentence from L_2.

Suppose now that we give to a person trained to translate from L_1 into L_2 the following text:

(t1) ‘C. B. A. C. A. B. C.’

Our translator will start with the first three sentences as expected, by writing '3. 2. 1.' She will nevertheless continue in a most strange way:

'1. 2. 3. 3.'

Confronted with the rule (R) she will answer that all the sentences were translated according to the rule, but she interpreted the phrase "replacing a sentence by another", as it appears in (R), in a different manner. She might express her interpretation like this:

(I1) For any two formal languages which both contain **m** sentences, replacing the $\mathbf{n}^{th}$ sentence from one language with the $\mathbf{n}^{th}$ sentence from the other means writing instead of the $\mathbf{n}^{th}$ sentence from the first language, alternately, once the $\mathbf{n}^{th}$ sentence from the second language and once the $\mathbf{\{n+1\}(mod\ m)}^{th}$ sentence from the second language.

In order to prevent such deviant interpretations we will make our intended interpretation explicit:

(I1') For any two formal languages which both contain **m** sentences, replacing the $\mathbf{n}^{th}$ sentence from one language with the $\mathbf{n}^{th}$ sentence from the other means writing the $\mathbf{n}^{th}$ sentence from the second language in the place of every apparition of the $\mathbf{n}^{th}$ sentence from the first language.

Now our translator will retranslate (t1) correctly:

'3. 2. 1. 3. 1. 2. 3.'

We make another try with the following text:

[6] Here ‘A’, ‘B’, ‘C’ and ‘1’, ‘2’, ‘3’ are not symbols or shortcuts for the “real” sentences of L_1 and L_2, but occurrences of the sentences themselves.

[7] Perhaps one could doubt that Wittgenstein would regard translating from L_1 to L_2 as a language-game. But then we could devise another example. We may consider, for instance, a situation where only three orders (‘A’, ‘B’ and ‘C’) are given and to understand them correctly is to react in a different way to each (one may think of transporting bricks from three different piles to a certain place, for instance). It is not important here that the reactions are not within the realm of linguistic behaviour anymore.

(t2) “C. B. C. A.”

This, however, produces an unexpected result: ‘1. 3. 2. 3.’ Again our translator claims that she agrees to the rule (R) *and* to the interpretation (I1'). Only that this time she has interpreted (I1') differently:

(I2) Writing an expression *in the place of* another which appears within one string of symbols is to create a mirror image of the string and substitute the expression accordingly.

Let us stop here for a bit and turn to some more general considerations[8]. For any two languages described as sets of sentences we could regard the translation rules as functions from the sentences of one set to the sentences of the other. The idea of correspondence between sentences must be somehow expressed in the rules. Some phrase marking the correspondence will appear in the formulation of the rule, playing the role the equality sign has in the definition of a function. This phrase could always be misinterpreted. We can imagine an entire class of non-standard interpretations of the translation rule in the following way:

For two finite sets of sentences having **n** elements, $S_1 \ldots S_n$ and $S'_1 \ldots S'_n$, there are $\mathbf{n^n}$ possible functions which associate elements from the second set to elements from the first set (of which, **n!** are bijections). Let $\mathbf{f_I}$ be the standard mapping function and **m** the number of possible functions. There are **m!** possible arrangements of the mapping functions in a row (**m!** is still a finite number). Let $\mathbf{f_1} \ldots \mathbf{f_m}$ be such a series. The pattern of a non-standard interpretation, then, could be:

(P) Up to the $\mathbf{i}^{th}$ occurrence of $\mathbf{S_k}$ the replacement should be made according to $\mathbf{f_1}$; from the $(\mathbf{i}+1)^{th}$ occurrence to the $(\mathbf{i}+\mathbf{j})^{th}$ the replacement should be made according to $\mathbf{f_2}$ etc.

Since **i**, **j** etc. are arbitrarily large numbers, it is obvious that there is an infinite number of possible interpretations of the translation rule[9]. Other patterns of non-standard interpretations are also possible, of course. There are non-standard interpretations of this sort, which will produce, in fact, the same result as $\mathbf{f_I}$[10]. So there is no point in saying that the standard translation rule should be privileged because it does not make the application of the function dependent

[8] In what follows I mainly expand on Wittgenstein, 1953 (from now on referred to as PU) §86 and §163.

[9] Actually, a procedure could be devised to assign to each instance of (P) a real number written in the base **n**, where **n** is an arbitrarily large natural number.

[10] Let the translation rule **R1** be that up to the 10^{th} occurrence of each sentence from L_1 ‘A’ should be replaced by ‘1’, ‘B’ by 2 and ‘C’ by ‘3’ and that the replacements should be made differently after that, no matter how. The translation rule **R2** will stipulate that up to the 10^{th} occurrence of each sentence from L_1 we should write ‘2’ for ‘A’, ‘3’ for ‘B’ and ‘1’ for ‘C’, but after that ‘A’ should be replaced by ‘1’, ‘B’ by 2 and ‘C’ by ‘3’. And now we can define **R3** by saying that up to the 10^{th} occurrence, each symbol from L_1 should be replaced by symbols from L_2 according to **R1**, and after that the replacement should be made according to **R2**. **R3** produces exactly the same results as (R), of course.

on the number of occurrences of the symbols which are to be processed[11]. There is no point either in trying to escape the difficulty by replacing the phrase expressing the correspondence relation in the formulation of the rule by the phrase used by the recalcitrant translator in *his* interpretation of the rule (the phrase "the replacement should be made" in (P), for instance). For, since the same phrase appearing in the context of different sentences could be interpreted differently, nothing prevents our translator to interpret the phrase differently, now that it appears within a different sentence from his. And there is, of course, no point in saying that it is not plausible that one would take "=" to mean something so different from what it means to us. Finally, specifying that "each occurrence of $\mathbf{S_k}$ must be translated in the same way" is useless, since "each" and "the same way" could be interpreted differently from what we intend them to mean[12].

It should have become clear by now that regardless of our stipulation of the translation rules, the correct interpretation of the rules, the correct interpretation of the interpretation of the rules and so on, the translator might always interpret all our explicit stipulations in a way such that both the translations which seem to agree with our intended rules and those which seem to break our intended rules can be made to agree with the explicit stipulations[13].

Perhaps one might want to reply now that our paradox appears only because we take the translator not to understand the metalanguage in which we formulate the translation rules. The metalanguage that the translator uses differs from ours in very few respects, though. Actually, the differences regard just some phrases used to express correspondence relations *in the context of the sentences stating the translation rules*. Apart from that, nothing is different. However, we cannot point to these differences properly without making appeal to the semantic rules that govern the use of those expressions in metalanguage. And here we only have an infinite regress. It is important to note, however, that the initial case had nothing to do with matters concerning meaning directly. The fact that our translator understands the meaning of 'A', 'B' 'C' and '1', '2', '3' bears no relevance on the example[14]. What we can say is that he does not understand "correctly translating from L_1 to L_2" as we understand it[15]. But I

---

[11] The source of this observation is Nelson Goodman's 'grue/bleen' example from Goodman, 1954.

[12] See PU § 185.

[13] Since **i**, **j** etc. may be arbitrarily large, the deviant interpretation of the rule can account for both the past translations which seemed to accord to the rule and for the present translations, performed after the $\mathbf{i}^{th}$ occurrence of the symbols to be translated (we could make interpretations dependent not on the number of occurrences of the symbols, but on the number of applications of the rule, too).

[14] We could even imagine that he will always substitute '1' for 'A', '2' for 'B' and '3' for 'C' when asked to write a sentence from L_2 which has the same meaning with a sentence from L_1. Nevertheless, he would not call this "translating from L_1 to L_2" (and neither should we).

[15] And this could be the case even if he *always* performs the translations as we do, for while we would be inclined to say that the rule we use is at least similar to (R), he might be guided by a rule similar to R3 (see note 7). A set of rules which produces the same result could be one specifying replacements for pairs of sentences (A.A. – 1.1., A.B. – 1.2., AC – 1.3., B.A. – 2.1., B.B. – 2.2. etc.). An extra rule will concern cases where we have an uneven number of sentences: at the end of the text to be translated, when we meet the last sentence, we add 'A' to it, translate the pair, and then erase the last '1' from the translated text (however, one

would like to suggest that, at least for now, we put matters concerning meaning and understanding apart and keep focussing on the relation between rules and their applications[16]. Our trouble, therefore, seems to be that it follows from the above example that we cannot distinguish between a correct translation and an incorrect translation by making appeal to the translation stipulations (rules, interpretations of the rules and the like).

Before going any further, there is one more point I wish to make. In the example above there are a limited number of cases that a translator has to deal with. This shows that our problems are in no way related to the fact that in applying a rule we might encounter new, unheard of, situations[17]. Nevertheless, each situation we would refer to as an "application of a rule" is, in a sense, a new case[18]. Only that this has not so much to do with the complexity of the activities (verbal or not) the rules regulate.

3. Our problems seem to arise when we try to bridge the gap between rules and their applications. We feel that a rule has to *determine*, so to speak, its correct applications in a necessary manner[19]. But insofar as we take the rule to be a linguistic entity, i.e. a sentence, we cannot argue for this strong feeling of ours. One might think that all our trouble comes from that fact that we took rules to be linguistic entities. If we could regard the rule as something which precludes the need for further interpretation, the paradox should disappear. Indeed, it could be said that a sentence could only be the expression of a rule and not the rule itself. The rule itself could be perhaps thought of as an abstract entity, which, by its own nature, contains its interpretation. All the cases in accordance with the rule are, somehow, determined by it. The problem with this view is that there is no apparent way in which we could trace the relation between the rule, seen as an abstract object, i.e. not existing in space and time[20], and the actual situations in which we say that someone has followed the rule. The cases in accordance with the rule could not be causally determined by the rule. There has to be another sort of relation. In order to show what this

---

might complain that this is not using a different rule, but only a different *procedure* to apply the same rule; this is disputable).

[16] Wittgenstein himself and most of his interpreters, most notably Kripke, seem to mix the question of the relation between a rule and its application with questions about meaning, understanding, the relation between the meaning of an expression and our linguistic behaviour exhibited in using it and so on (but see PU §§156-178). If the rules concerned are semantic rules, this is, I think, fully justified. However, the extent of Wittgenstein's RFC goes far beyond that. In this I agree with Robert Brandom (see Brandom, 1994, p. 15). The rules I am mostly concerned with govern the distinction between meaningful and meaningless sentences, but have nothing to do with establishing the meaning of sentences or their parts (since they are supposed to be prior to any proper semantical matters).

[17] Wittgenstein's own examples of learning how to write down or to continue a series of numbers and Kripke's example about addition seem to suggest that. One can also develop an argument going this way from PU §80, for instance, but this is incidental to our discussion.

[18] The extent to which the context is making a contribution to a case is best presented in Wittgenstein, 1956, VI-34 par. 3-4.

[19] See, for instance, PU §188 par.2.

[20] This also means that we cannot point directly to the rule and say to our recalcitrant translator: "This is the rule you were supposed to obey".

relation consists in, we must talk of some correspondence. And here a version of Plato's "third man" argument will surely apply.

Another strategy would be to speak of the rule as a mental object. The problem with our recalcitrant translator seems to be, on this view, that he has not grasped the translation rule yet. The rule is not "in his mind", so to speak. Were the rule "in his mind", there would be no more questions of interpreting or applying it wrongly. What the above example shows is at best that we cannot "put the rule in the mind of the translator" by using rational arguments. This does not make using rational arguments unnecessary, of course. We could show to someone who has already grasped the rule that he had broken it in one particular case by using such arguments. Let us ask now what sort of mental object the rule is. Suppose it is a mental representation. For the case presented above, grasping the rule might be, for instance, having the representation of a translation table. The table will perhaps look like this:

A $\longrightarrow$ 1
B $\longrightarrow$ 2
C $\longrightarrow$ 3

We still have the problem of using the table in actual cases. The three arrows could be regarded as a *single* symbol, which provides a substitution rule for 'A', 'B', 'C' and '1', '2', '3'[21]. As a *single* symbol, it may stand for *any* substitution rule, of course. One might try to answer to this problem by saying that we actually have three different representations, one for each row of the table, which are kept apart from each other in our mind. I will put aside for now the objection that if the three representations are kept apart from each other there is no indication that they have something in common, which they should, since they relate to the same rule. Let's accept for now that things are this way. In order to prevent further misinterpretations, we could even eliminate the arrows. The three representations will then look like this:

A1
B2
C3

One problem with 'A1', 'B2' and 'C3' is that they contain no hint that they have anything to do with a rule. But even if we accept that the person who represents 'A1' to herself somehow knows that 'A1' is a rule, she still might get the rule wrong. There are many ways in which one can make substitutions while having in mind the three representations from above. The string 'A. C. B. C', for instance, could produce any of the following results:

'A. CCC. BB. CCC.'
'1. 333. 22. 333.'
'A1. C3. B2. C3.'
'A. B. C. C.'[22]

[21] See, for instance, PU § 60 for the idea that there is no absolute distinction between simples and complexes.

[22] This result is explained by the fact that the translator takes 'A1', 'B2' and 'C3' to mean something like a sorting rule: "Rewrite the series of symbols such that, in the end, all the A's

‘1CBC. A3B3. AC2C’
etc.

We expected the translator to read ‘A’, ‘B’, ‘C’ and ‘1’, ‘2’, ‘3’ appearing in his mental representations as iconic symbols for the signs she has to work with. It is obvious that she did not do so and is not at all clear why she *should* have done that. And here there still is a correspondence relation we seem to have overlooked, namely that between the elements of our representation (‘A’, ‘1’, ‘2’ as images in our mind) and the components of the cases to which our representations ought to apply (‘A’, ‘1’, ‘2’ written on a piece of paper).

This amounts to the conclusion that grasping the rule is not having a mental representation. Let us now substitute the concept of “a sign looking like ‘A’” for the mental representation of ‘A’, the concept of “a sign looking like ‘1’” for the mental representation of ‘1’ and so on. The person who has grasped the rule that ‘A’ should be replaced by ‘1’, then, must possess, apart from mental representations of ‘A’ and ‘1’, the concept of a sign which looks like ‘A’ and the concept of a sign which looks like ‘1’. She must also possess the concept of “substituting one sign for another”. And now the rule could be regarded as a proposition formulated in the language of thought:

(R-LOT) Substitute the sign, which looks like ‘1’ for the sign which looks like ‘A’.

The fact that this is a thought and she possesses the right concepts, we think, prevents any misinterpretations of the rule. But here we are in no better position than when we spoke of the rule as an abstract object. Neither an abstract object, nor a mental one has feelers to touch the reality with[23]. This is perhaps inherent to the way we usually speak of objects and relations: one object does not contain the relation it has with another.

4. Let us now think that we have a large enough set of cards. Some of them have ‘A’ written on one side and ‘1’ on the other. Others have ‘B’ written on one side and ‘2’ on the other. Those with ‘C’ written on one side will display ‘3’ when turned over. A text written in L_1 by using the cards could be correctly translated into L_2 by turning all the cards to the other side. I will further assume that this is done mechanically. Perhaps we could say that the mechanism formed by the cards and the means to turn them over is correctly translating from L_1 to L_2. The translation rules are somehow physically instantiated in this mechanism, and the relation between the rules and their correct applications is a causal relation. And now it may occur to us that we are not in a very different position. Our brains are, in a sense, such mechanisms, though far more complicated. This is the starting point of a naturalist approach. The problem with our puzzle, as it may occur to the naturalist, was due to the fact that we were trying to get from rules to their correct applications by using a “logical path”, so to speak. This choice was enforced upon us by our preference for the method of conceptual analysis. Yet, this was not the only available alternative. Another way of getting from the translation rule of our example to the set of appropriate translations would be that of stating the

will come first, all the B’s will come in the second place and all the C’s in the third.”

[23] The earlier Wittgenstein thought that of pictures (see Wittgenstein, 1961, 2.1515).

causal laws which describe the functioning of the mechanism translating from L_1 into L_2[24].

The approach proceeds with the remark that grasping a rule is some sort of ability or disposition[25]. However, it will be conceded that when we speak of dispositions or abilities we actually speak of certain processes going on in our brains, processes which cause us to behave as we do when we act according to the rules. Explaining the relation between rules and acting according to rules is nothing else but constructing an empirical theory about those processes and the relations between them and our behaviour.

Let us now try to figure out how a theory explaining the relation between (R) and correctly translating from L_1 to L_2 would look like. One model for the facts the theory must apply to is this:

Translation Input ⟶ Translating Mechanism ⟶ Translated Output

The theory, then, will include some general principles or laws that govern the functioning of the Translating Mechanism. From these laws, with the aid of some derivation principles, we should be able to obtain predictions of the form:

(P) Under the circumstances C_1&…&C_n, if state IS obtains in the mechanism, then state OS will also occur.

'IS' and 'OS' are two states of the Translating Mechanism, which correspond to a Translation Input and to a Translated Output respectively[26]. Here it is clear that our example helps us to keep things simple: there is no need to speak of an infinite number of states occurring in the mechanism[27]. There are three IS states and three OS states. The C_1,…,C_n conditions are required in order to "cut the causal relations" between the functioning of the mechanism and everything else going on in the universe (my head included)[28].

---

[24] See, for instance, Pylyshyn, 1998 for this view.

[25] See the first two sentences of PU §149.

[26] One might wish the theory to contain also an explanation of the causal relation between the input and the output facts (i.e. signs written on paper, for instance) and the IS and OS states. That part should explain, for instance, what happens when the translator mistakenly takes a 'B' in handwriting for a 'C' and translates it by '3'. However, I am not concerned with this sort of situations here.

[27] See note 15. The rule of addition or the rule of writing the series of even numbers still apply for numbers so large that we would never be able to represent, given the finite number of neurons in our brains. But I do not think that *this* counts as a strong critique of the naturalist view (It would be very easy to speak of rules for performing addition or obtaining the next even number for numbers written in unary base. And we could apply these rules even to numbers that we do not read entirely.)

[28] Is our theory supposed to account for neural processes that correspond to "willing to translate correctly from L_1 to L_2"? I think not. Yet, it is obvious that even if the ability to translate correctly is instantiated in my brain and the IS corresponding to reading 'A' obtains, I could write down '3' just because I *do not want* to translate 'A' correctly. So, one of the conditions should specify that I want to perform the translations correctly (i.e. that the state corresponding to that occurs in my brain, perhaps in some other region than the one where the translating module is located).

There is a question that could be raised now: How general are the laws of the theory? We could think of many different mechanisms, apart from our brains, performing what we would call "translations from L_1 to L_2". We could use water pipes, strings, electric wires etc. in order to build such mechanisms. The Translation Input will be coded differently, for each such mechanism, of course. If we build a different theory for each, then we should say that there are different rules for "translating from L_1 to L_2 with a computer", "translating by using a human brain" and so on. On the other hand, it is obvious that the functioning of a machine using water is governed by different physical laws from those concerning the functioning of an electrical machine. The idea of *coding* the input and the output brings some problems too: we could think of exactly the same states of a mechanism coding a different input or output. For instance, let L_3 be a language which contains only three sentences: 'I', 'II', and 'III'. It is clear that exactly the same theory that explains the functioning of our first mechanism using cards to perform the translation will work as an explanation of a translation from L_3 to L_2. Should we say that we only have one translation rule or two? I will suppose, for the sake of the argument, that the laws of our theory apply only to human brains[29] and that we could somehow circumvent the idea of coding by speaking of the causal relation between the input and the output, on one hand, and the brain states, on the other. Thus, the occurrences of the sentences from L_3 will produce different states in our brain from those produced by the occurrences of the sentences from L_1. But we might still feel that there is something not completely in order here. How is it possible that we could derive only three predictions (connecting the state of reading 'A' with the state of writing down '1', the state of reading 'B' with that of writing down '2', and the state of reading 'C' with the state of writing down '3') from our theory? Were the languages L_1 and L_2 one sentence bigger, would we have to construct a completely different theory? To this one could answer that perhaps there are two kind of laws or principles: some general principles which will perhaps explain all the cases of the 'writing something instead of something else' kind, and some particular laws which, when added to those general principles, produce only the three predictions we were looking for[30].

Suppose now that we have completed this theory. Its structure could be roughly represented like this:

$$T = \langle GL, PL, DP, M \rangle$$

where GL is the set of general laws of the theory, PL is the set of its particular laws, DP is the set of derivation principles and M is the meta-theoretical assertion that the theory offers an explanation for what is to perform a correct translation from L_1 to L_2. Maybe one would like to say that the rule (R) was replaced in this theory by the laws from the PL set. This, I think, is not a very good move, since it instantly brings forward questions related to the distinction

[29] One could say, for instance, that the theory regards "consciously translating from L_1 to L_2", although it does not necessarily have to explain the part regarding consciousness. Being conscious about what one is doing could be regarded as one of the conditions $C_1,\ldots,C_n$ and the explanation for that could be left as a task to a different theory.

[30] Someone might think here of problems like: 'Is the functioning of the brain of a person who knows to replace symbols and learns to translate from L_1 to L_2 similar to that of the brain of a person who learns to replace symbols *by learning* to translate from L_1 to L_2?' (compare this with PU §§ 156-158)

between norms and natural laws. So perhaps is better not to try to pin the rules on any assertion of the theory[31]. The consistent naturalist will probably say that it is not the case to think of the rule as a linguistic entity anymore and therefore it makes not sense to speak of a distinction between two types of statements. That is a fair enough remark, I think.

Yet, there might be a problem with this naturalist account[32]. Since T is an empirical theory, it should be at least in principle refutable. Let us try to find out what would count as an empirical infirmation of T. Let us think of a translator who translates correctly from L_1 to L_2. The states described by the theory occur in his brain, according to T's laws and all, the conditions $C_1,\ldots,C_n$ are satisfied and the IS corresponding to reading 'A' also obtains. Yet he makes something which we would usually call a mistake. This happens perhaps due to a disturbing factor which the theory did not account for. There are two different ways in which we could modify our theory to fit the empirical facts. One is to add the condition that the disturbing factor does not occur to the conditions $C_1,\ldots,C_n$. Another possibility is to extend the laws of the theory such as to cover these cases and predict the translator reactions in these situations too. Only that in this situation our theory will not be a theory of "correctly translating from L_1 to L_2" anymore[33]. But why do we say that? This was not an infirmation of M. Since a correct translation is whatever the theory describes, why not adopt the second choice? And if we do not adopt it, is it not because we make appeal to the rule?

Let us now imagine a different situation. Our translator makes the same mistake. And now we realise that this too can be accounted for by the theory. We have thought that there are only three predictions that could be derived from T, but we applied the principles from the DP set incorrectly. And we realise that the translator's behaviour could be also predicted by our theory. So there is no empirical infirmation here. We have discovered that the explanatory power of our theory is greater than what we have thought. Still, we wish to change our theory. Why is that so?

Perhaps our problem can be pointed at directly in the following way. If T is to be an empirical theory, then M too, since it is a part of it, must be refutable by the experience[34]. But how can we speak of an empirical infirmation of M without making appeal to the non-naturalised concept of the translation rule? Being a meta-theoretical statement, M states a relation between the rest of the theory and the non-naturalised rule. It is obvious that we cannot give M up. But what can we make of it?

---

[31] And neither on the theory's predictions. Indeed, it is quite easy to see that stating a causal connection between reading 'A' and writing '1' is no replacement for a translation rule, seen as a norm, since speaking of what will or what would happen, were some other things to happen too, is completely different from speaking of what *ought* to happen. One might say, for instance, "Even if you are brain-dead, it still applies that you *must* translate 'A' by '1'."

[32] In what follows, the background of my remarks is contained in PU §158 and PU §§ 193-195.

[33] It could be, perhaps, the theory of "translating correctly from L_1 to L_2 and making mistakes when you are very tired".

[34] This does not mean that it should be refutable considered in isolation from other principles of the theory. Actually, because M makes reference to all the other statements of the theory, it is only in conjunction with *all* of them that its sense could be rendered properly.

The naturalist's reply, as I imagine it, could run as follows: "What guides our investigation is the model of a mechanism which never makes mistakes. Our theories are empirical in the sense that they are only approximations of such a model. The more factors which could affect the correct functioning of the translating mechanism we discover and eliminate, the better our approximations of the ideal model are. So M states the relation between our theory and this ideal model. We don't make appeal to a non-naturalised rule, but to an ideal model within which the causal relation between the processes instantiating the rule and those counting as applications of the rule is never affected by disturbing factors." This account, however, seems to contradict our common view about the scientific progress. We use to think that science advances by offering models that are less ideal and more close to the facts[35]. In this vein, one may think that Classical Mechanics is a more abstract model of the physical interactions than the Relativity Theory, which offers a more accurate description of the facts[36]. Why would anyone want to go the other way round? But I will leave that aside for now. Let us imagine that we have two such ideal models. One is that of a regular translator who never makes mistakes. The other one is a model of one of our deviant translators who never diverges from the way he is performing the translation. Why do we choose to be guided by the first and not by the last model? How could the naturalist motivate his choice of the first model without speaking of the translation rule? And even if he could somehow motivate his choice, there still is a problem left. On the new reading, what M says is that there is a relation between T\{M} and the ideal model. Since the ideal model is not a physical object, it follows that the relation cannot be a causal one, which makes it obvious that M could not be empirically infirmed. If the naturalist is not bothered by that, it is only because he does not realise that now he is in exactly the same position as the conceptual analyst. Namely, if he wants to hold on M, he must speak of the relation between an abstract object (the ideal model) and actual situations and processes. If he only wants to speak of the functioning of our brain, then nothing will be objected to that. But then he will not be in a position to claim that he has answered a naturalised version of the question regarding the relation between rules and their applications.

5. If it is true that the both adumbrated approaches fail to account for the relation between rules and their applications, then it is hard to see how a mixture of the two will succeed to do the job. And it is useless, of course, to try to state rules if we cannot relate them to their applications properly. What is wrong with (A), then? Perhaps it is wrong to speak of rules and their application as if they were two different things. Anyway, I am not going to try to advance a solution to our problem here but only to evaluate the import of the arguments presented above.

To prevent any misunderstandings, it does not follow from the above presentation that we cannot communicate, compute, that our social institutions

[35] I think Max Planck was expressing this pre-kuhnian view very nicely. And even if this view is perhaps obsolete, I do not see how bringing a more sophisticated account of the scientific progress could change the odds in the favour of the naturalist.

[36] In taking this to be our common view about the relation between these two theories I deliberately disregard more learned accounts of this matter like those offered by Einstein and Heissenberg. I do not think, however, that this vitiates my argument.

are illusory or anything of the kind. Nor do I think that any relativist conclusions are supported by this account alone. What the arguments from above amount to is only that we cannot produce normative *theories*[37]. Again, this is not to say that there are no values, norms or rules. They can be taught, transmitted, even enforced. They sometimes collide. There is nothing wrong with normative talk either, as long as we do not take normative talk for what it is not, namely a descriptive account of norms, rules and values[38]. Rejection of theorizing about norms and values might be taken by some to lead to some sort of irrationalism. Since we cannot provide theories and arguments about what is morally, prudentially, practically or in some other way reasonable to do in this or that situation, one may think, it is as if we have no reasons at all. I do not agree with this view. Discussing these matters is, however, far beyond the topic of my paper.

*

It seems quite radical, to say the least, to maintain, as Wittgenstein usually does, that all the traditional philosophical questions are meaningless. It is difficult, nevertheless, to remove the suspicion that some of them might actually be nothing more than nonsensical puzzles. We do not have solid, ultimate criteria to distinguish between meaningful and meaningless sentences. And it might be the case that is in principle not possible to have such criteria. Trying to discover by philosophical analysis the rules that govern the distinction and to state them is doomed to be a Sisyphean activity. Constructing scientific theories which explain what happens in our brains when we learn to use the language or when we actually distinguish between meaningful and meaningless sentences could be informative, but will not help us a single bit to remove the shadow of suspicion cast upon our philosophical activities. Most philosophers, I do hope, will not think that ignoring Kant and plunging into traditional metaphysics would be very wise. And some would perhaps agree that ignoring the later Wittgenstein or interpreting him such that his views do not endanger our philosophical habits is not advisable either.

[37] This is the conclusion of Wittgenstein, 1961, of course. I don't think that the later Wittgenstein has changed his view in *this* respect.

[38] There are cases in which we would say that normative talk is inappropriate even if no one would take it for theorizing. We might feel, for instance, that a sentence like "Since you are a very close friend of mine, you ought to express condolences to me for the death of my mother" should *never* be uttered.

References

Boghossian, Paul, 1989, "The rule-following considerations", *Mind*, pp. 507-549

Brandom, Robert, 1994, **Making it Explicit**, Cambridge, MA: Harvard University Press

Evans, Gareth, 1985, **Colected Papers**, Oxford: Clarendon Press, pp. 322-342, "Semantic Theory and Tacit Knowledge"

Goodman, Nelson, 1954, **Fact, Fiction, and Forecast**, Cambridge, MA: Harvard University Press

Hacker ,P. M. S. and Baker, G. P., 1984, **Language, Sense and Nonsense**, Oxford: Blackwell

Hacker, P. M. S. and Baker, G. P., 1985, **Rules, Grammar and Necessity: An Analytical Commentary on the Philosophical Investigations**, Oxford: Blackwell

Kripke, Saul, 1982, **Wittgenstein on rules and private language**, Cambridge: Harvard University Press

McDowell, John, 1992, "Meaning and Intentionality in Wittgenstein's Later Philosophy", *Midwest Studies in Philosophy*, vol. XVII, pp. 40-52

McDowell, John, 1998, **Mind, Value and Reality**, Cambridge, Massachusetts & London: Harvard University Press, pp. 221-262

Pylyshyn, Z., 1998, article on "Cognitive architecture" in Routledge Encyclopedia of Philosophy, London: Routledge

Wittgenstein, L., 1953, **Philosophical Investigations**, edited by G. H. von Wright, R. Rhees, G. E. M. Anscombe, translated by G. E. M. Anscombe, Oxford: Basil Blackwell

Wittgenstein, L., 1956, **Remarks on The Foundation of Mathematics** *(Bemerkungen über die Grundlagen der Matematik)***, e**dited by G. H. von Wright, R. Rhees, G. E. M. Anscombe, translated by G. E. M. Anscombe, Oxford: Basil Blackwell

Wittgenstein, L., 1961, **Tractatus logico-philosophicus**, London: Routledge & Kegan Paul

Wright, Crispin, 1981, "Rule-Following, Objectivity and the Theory of Meaning", in Holtzman, Steven H. & Leich, Christopher M. (eds.), **Wittgenstein: to Follow a Rule**, London: Routledge & Kegan Paul, pp. 99-117.