

CDD: 149.9

WHY FREGE THOUGHT IT TO BE “PROBABLE” THAT TRUTH IS INDEFINABLE*

MARKUS STEPANIANS

Department of Philosophy
Universität des Saarlandes
FR 3.1, D-66041 Saarbrücken
GERMANY

stepanians@compuserve.com

Abstract: Frege’s so-called “Regress Argument” is closely examined and it is argued that Dummett’s reconstruction of it is not satisfactory. *Contra* Dummett, the argument does not involve a regress, is not a *reductio* and not even a strictly deductive argument. Rather, what Frege tries to show is that any attempt to define truth fails to be epistemically fruitful and thus misses the very point of analytic definitions of concepts. The cause of this epistemic sterility is an inherent circularity, and it is in view of this defect that Frege suggests that truth is very likely so simple that it cannot be defined at all. This circularity, in turn, is due to the fact that Frege takes the sense of “true” to be a constitutive part of the content of acts of thinking a complete thought. Finally, I trace Frege’s belief in this

*An earlier version of this paper was presented at the Frege Workshop in Leiden, August 23rd – 25th, 2001. I am very grateful to the organizers, Göran Sundholm and Kai Wehmeier for their invitation, and to the participants of the workshop and for valuable suggestions, especially Jamie Tappenden. I would like to thank David Wiggins for his kind encouragement and very welcome corrections of my English. My main debt, however, is to Ulrich Pardey, whose extensive and probing comments on a condensed version of the argument as presented in Stepanians, 2001, chap. 8, §§1-3, forced me to clarify and develop my thoughts on this topic further. For Pardey’s views, see Pardey, 2004.

Manuscripto – Rev. Int. Fil., Campinas, v. 26, n. 2, pp. 331-345, jul.-dez. 2003.

omnipresence of the sense of the word “true” in thought to his conception of judgment and suggest that it may very well be false.

Key words: Frege. Truth. Thought. Judgment.

In “Thoughts” Frege offers a notorious piece of reasoning that is supposed to show not only that it is impossible to explain truth in terms of correspondence but that any attempt to define truth is likely to fail. Suppose, says Frege, we defined truth in terms of correspondence:

But then what ought we to do in order to decide that something is true? We would have to investigate whether it would be true that – say, an idea and something real – coincided in the respect thus determined. And with that we would be confronted with a question of the same kind and the game could start from the beginning. So this attempt to explain truth as correspondence fails. But so any other attempt to define truth fails, too. Because in every definition one would have to specify certain marks. And in an application to a special case what would then matter is whether it would be true that these marks applied. So one would go round in a circle. Hence it is probable that the content of the word ‘true’ is unique and indefinable. (*Ged*, 60; my transl.)¹

Twenty five years ago, Michael Dummett gave a famous reconstruction of this “Regress Argument”, as it has been called since, and argued that it fails to establish its intended conclusion (cf. *FPL*, 443).

¹ *“Was müßten wir dann aber tun, um zu entscheiden, daß etwas wahr wäre? Wir müßten untersuchen, ob es wahr wäre, daß – etwa eine Vorstellung und ein Wirkliches – in der festgesetzten Hinsicht übereinstimmen. Und damit ständen wir wieder vor einer Frage derselben Art, und das Spiel könnte von vorn beginnen. So scheitert dieser Versuch, die Wahrheit als eine Übereinstimmung zu erklären. So scheitert aber auch jeder andere Versuch, das Wahrsein zu definieren. Denn in einer Definition gäbe man gewisse Merkmale an. Und bei der Anwendung auf einen besonderen Fall käme es dann immer darauf an, ob es wahr wäre, daß diese Merkmale zuträfen. So drehte man sich im Kreise. Hiernach ist es wahrscheinlich, daß der Inhalt des Wortes, wahr’ ganz einzigartig und undefinierbar ist.”*

Although Dummett’s negative verdict seems to be widely accepted, this has not stopped philosophers from thinking and writing about Frege’s argument. To many it seems that, even if Frege’s reasoning fails as an argument for the indefinability of truth, it *does* show something of great importance – if only we knew what. Several candidates for the “proper” conclusion of the Regress Argument have since been offered. It has been taken to deliver a final blow to the correspondence theory of truth (Dummett); to prove that a theory of truth must allow for the spontaneity of judgment (Bell); it has been thought to be destructive of any empirical theories of truth (Carruthers); and recently: to prove that truth is not a real property (Ricketts).

Whatever the Regress Argument as stated by Dummett might show, I do not think it captures Frege’s original intentions. In what follows I will first question Dummett’s reconstruction and then suggest an alternative. I think that the argument as intended by Frege does not involve a regress at all, that it is not a *reductio* and not even a strictly deductive argument with the indefinability thesis as its conclusion. Rather, what Frege tries to show is that any attempt to define truth fails to be epistemically fruitful and thus misses the very point of analytic definitions of concepts. That any definition of truth is, as a matter of principle, epistemically sterile is the main conclusion of Frege’s argument. The cause of this epistemic sterility is an inherent circularity, and it is in view of this defect that Frege suggests, in a second step, that truth is very likely (“probable” [*wahrscheinlich*]) so simple that it can not be defined at all. I will suggest that this circularity, in turn, is due to the fact that Frege takes the sense of “true” to be a constitutive part of the content of acts of thinking a complete thought. To grasp the thought that *p* is to think that it is true that *p* (although it does not, of course, imply *judging* that it is true). Finally, I trace Frege’s belief in this omnipresence of the sense of

the word “true” in thought to his conception of judgment and suggest that it might very well be false.

1

According to Dummett, Frege intended a *reductio* along the following lines: 1. The attempt to apply a definition of truth leads to an infinite regress. 2. But a definition that it is impossible to apply is absurd. Therefore, 3. truth is indefinable. The crucial question, then, is this: What gets the regress going? Which feature of the definition triggers it? It seems to be uncontroversial that the decisive point in Frege’s argument is the application of the marks given in the definition to a special case. As he says in the text I quoted at the beginning:

Because in every definition one would have to specify certain marks. And in an application to a special case what would then matter is whether it would be true that these marks applied. So one would go round in a circle.²

Let the concept F be one of the marks specified in the definiens. Then to decide whether x is true, says Frege, we should have to decide whether it is true that it is F .

That seems correct. But why is it a problem? According to Dummett, it is a problem because it triggers an infinite regress. In order to apply our definition, we have to ask ourselves whether x is F ; but in order to answer this question, we would have to decide *first* whether it is true that x is F ; but the answer to that question presupposes an answer to the question whether it is true that it is true that x is F – and so on *ad*

² “Denn in einer Definition gäbe man gewisse Merkmale an. Und bei der Anwendung auf einen besonderen Fall käme es dann immer darauf an, ob es wahr wäre, daß diese Merkmale zuträfen. So drehte man sich im Kreise.”

infinitem. Each judgment in this infinite series presupposes another, different judgment and we are never in a position to judge anything. So according to Dummett, the implicit assumption that gets the regress going is this:

(R) To judge that x is F one has *first* to decide whether it is true that x is F .

If we make the appropriate substitutions it becomes clear that (R) indeed leads to an infinite regress. And so (R) seems to be the crucial premise in Frege's argument.

But this is puzzling. First of all, as Dummett himself says, it is not hard to see that (R) is absurd. If true, it would make judging itself impossible. It would not only block applications of the marks given in a truth definition but applications of any concept whatsoever, and would thus frustrate all judgments. Moreover, (R) is not only absurd, Frege himself would be the first to point this out. For him, to wonder whether something is F is to wonder whether it is true that it is F . In his view, what we have here are two descriptions of *one and the same* mental act, whereas (R) implies that we are dealing with two different acts that can be executed one after the other. But if this is so, why should Frege invoke a premise that he believes to be obviously false? This seems odd. Could it be that Frege believed that any attempt to define truth somehow involves a commitment to (R)? But this also does not seem plausible. Not only is it difficult to see why even the attempt to define truth should imply (R), there is no trace of such an argument in Frege's text. And in any case it would be a quite different argument from the one Dummett identified, since (R) would not be another premise, but already the absurd conclusion of Frege's alleged *reductio*.

2

We have reached an impasse that justifies a fresh look at Frege's text. On closer inspection we will find, I think, that he neither here nor anywhere else in his writings appeals to (R), not even in the course of a *reductio*. Quite to the contrary: When Frege says that the application of F to x necessarily raises the question of whether F is true of x , there is no suggestion that one has *first* to do one thing before being in a position to do another. On the contrary, as I already said, he takes for granted that we are dealing here with one and the same mental act. Far from invoking (R), Frege is relying on something incompatible with it, namely the identity:

(I) Thinking (wondering, assuming, judging, etc.) that x is F =
 thinking (wondering, assuming, judging, etc.) that it is true that x is
 F .

But, if (I) rather than (R) is the decisive premise in Frege's argument, and (R) is responsible for the regress, it becomes doubtful that it does involve a regress at all. Since according to (I), the act of thinking that x is F , and the act of thinking that it is true that x is F , and the act of thinking that it is true that it is true that x is F , etc. are strictly identical, there is no suggestion of an infinite series of acts, not even of a finite one. There is no series at all, no multitude of acts but only *one* act, and hence no progression, no regress. The impression to the contrary is due to the possibility of *describing* this very same act in different, indeed infinitely many ways. This multitude of descriptions suggests an actual progression where we have just an endless repetition of the very same act.³

³ Indeed, Dummett's reconstruction of Frege's reasoning illustrates the dangers of speaking of *sentences* as being true or false in this context. He writes:

In an earlier text, Frege illustrates his argument by appealing to the picture of "a man on a treadmill who makes a step forward and upwards, but the step he treads on keeps giving way and falls back to where he was before" (*PW*, 134). There is a subtle, but significant difference between Frege's man on the treadmill and the man who, under the spell of (R) as suggested by Dummett, has to run *per impossibile* through an infinite series of judgments. Dummett's eternal runner is someone who – like Achilles chasing the Tortoise – can never reach his goal, but he is constantly making at least *some* progress. Achilles can not catch up with the Tortoise, although he does come closer to her all the time. But Frege's man on the treadmill just repeats the first step all over again without getting anywhere – just like someone who judges that x is F , that it is true that x is F , that it is true that it is true that x is F , and so on – exactly as we should expect in the light of Frege's principle (I). Although in "Thoughts" Frege does not mention the man on the treadmill, the picture of "going around in a circle" he invokes there also supports the view that (I) is at work in the argument, not (R). The upshot is the same in both metaphors. At the end of the day, the man on the treadmill and the man going in circles are both – unlike Achilles – exactly where they departed. Moreover, although the man on the treadmill is in a rather sad position, there is no absurdity involved. To reach the conclusion

"The same reasoning shows that truth is absolutely indefinable: for, if the truth of a sentence were to be defined as its possessing such-and-such characteristics, we should have, in order to determine whether the sentence was true, to enquire into the truth of the sentence which ascribed those characteristics to the first sentence; and again we should be launched on an infinite regress" (*FLP*, 443). What seems to have misled Dummett is that there are *two* sentences, but only *one* thought: The "sentence which ascribed those characteristics to the first sentence" is indeed different from the one it is about – but the thought expressed is the same in both cases. The regress exists only at the level of sentences, not of thoughts.

that attempts to define truth end up in that frustrating situation is certainly unfortunate, but not absurd. This suggests that Frege's argument does not involve, pace Dummett *et al.*, a real *reductio*.

It is time to put these illustrations and metaphors aside and try to get clear about what they are supposed to illustrate. What is the goal that cannot be reached, the intention that is necessarily frustrated in any attempt to define truth? Well, a proper truth definition, of course – but what does that imply for Frege? In order to answer this question we have first to understand the general point of the activity of defining concepts. Since what we are after when we ask for a definition of truth is an explanation of a concept already in use, it is obvious that Frege is denying the possibility of analytical [“*zerlegendè*”] as opposed to stipulative [“*aufbauendè*”] definitions of truth (for this distinction, cf. *PW*, 207ff). But first we have to be clear about the point of analytical definitions. This will enable us to see what can go wrong with them.

Before I adress this question, let me summarize what I have said so far. Dummett's reconstruction is on the wrong track. First, there is no implicit premise (R) at work in Frege's argument, but the principle (I) which is incompatible with (R). Second, there is no regress, but a kind of circularity. And finally, Frege's Argument does not have the form of a *reductio*. He is not trying to show that attempts to define truth lead to absurd conclusions. As we shall see in a moment, they simply fail to achieve their intended purpose. But they do so for reasons of principle, and this suggests to Frege that truth might be undefinable.

3

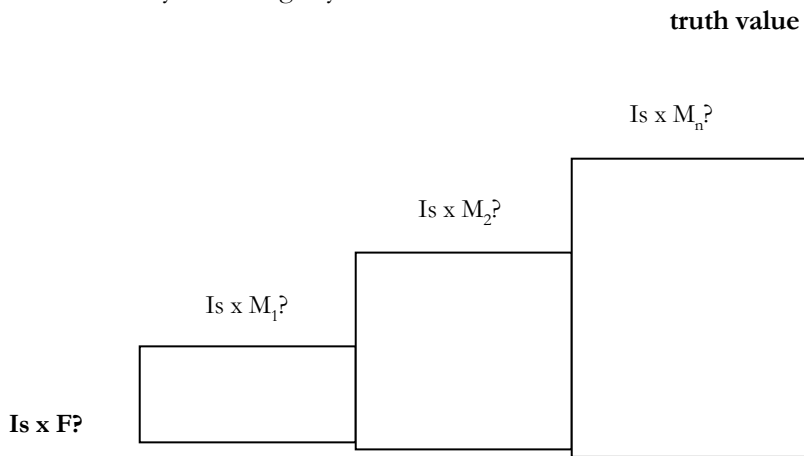
Everything we do in science is ultimately done for the sake of knowledge, and this goal is also what gives the activity of analysing concepts its point. For Frege, a good analytical definition is fruitful in the

sense that it helps to decide questions we would be unable to decide without it. And it is here that definitions of truth fail to deliver for reasons of principle. If this is correct, the frustration that awaits anyone who tries to define truth is of an epistemic nature and I suggest we call this defect "inherent epistemic sterility". We now have to answer two questions. First, how do fruitful analytical definitions help to achieve new knowledge? And second, what is it about definitions of truth that is responsible for their inherent epistemic sterility?

By giving an analytic definition, says Frege, "we carry out a logical analysis of a logical structure in order to find out what its constituents are" (*PW*, 208 f.). As a result, we might be able to solve problems we were unable to solve before. In *Logic and Mathematics*, Frege says, "it may not be possible to prove a truth containing a complex constituent so long as that constituent remains unanalysed; but it may be possible, given an analysis, to prove it from truths in which the elements of the analysis occur" (*PW*, 209). Frege might be thinking here of his own attempts to provide analytical definitions of concepts like *natural number* and *successor*. But the same point can be made for scientific research in general. The successful analysis of key concepts like *knowledge*, *logical inference* or *human right* will not only give many questions in Epistemology, Logic and Moral Philosophy a precise sense, it will often allow us to decide difficult cases as to whether what we have before us is really knowledge, is really an inference or a human right. Suppose that knowledge is justified true belief. The point of such an explanation lies precisely in the fact that it permits us to translate the question whether *NN* knows that *p* into a series of different questions about the marks specified in the definiens: Does *NN* believe that *p*? Does he have a justification for it? And finally: Is it true that *p*? What is important is that these questions can be answered without mastery of the concept of knowledge. Even someone who does not yet have that concept in his repertoire could answer them,

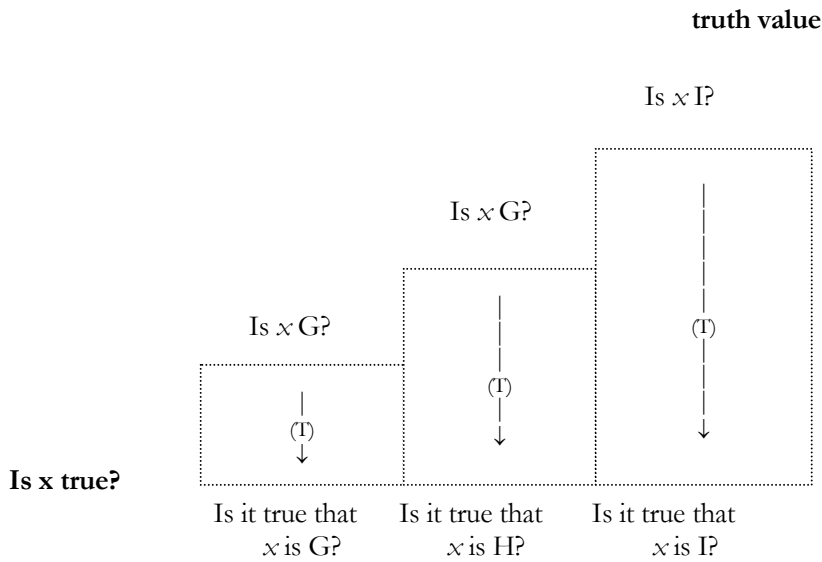
at least in principle. And that's why the definition could be used to teach someone that concept.

With the help of Frege's metaphorical characterization of judging as "advancing from the thought to its truth value", we might put the same point this way. The marks $M_1 \dots M_n$ given in the analysis of a concept F allow us to partition, so to speak, the total distance between the thought that x is F and its truth value into "smaller" ones that we can tackle independently. Every time we answer one of the questions concerning one of the marks specified in the definiens we climb one step upwards, make some progress and advance a bit further towards the truth value of the thought expressed in our original question. And even if we can not answer all of them, there is a clear sense of moving closer to the solution by answering any:



This is the general model for fruitful definitions. The reason it does not work for truth definitions is Frege's principle (I). Suppose we proposed to define truth in terms of the concepts G , H , and I . This would allow us to translate the question whether x is true into three

questions: $Is\ x\ G?$, $Is\ x\ H?$, and $Is\ x\ I?$ But unlike in the case of our knowledge example, none of these three is independent of the original truth question. In virtue of (I), to wonder whether x is G is to ask a truth question, namely whether it is *true* that x is G , and so for all of them. Thus, as Frege says, we have here "a question of the same kind" to the one we had hoped we could at least partially answer with the help of our analytical definition, namely a truth question. Instead of climbing upwards one step and advancing a bit further towards the truth value of our original thought, we are pushed down again to ground level, as it were. We are in the frustrating situation of Frege's man on the treadmill:



Noting that the notion of truth "pops up", so to speak, whenever we try to reduce it to simpler concepts, Frege declares it to be "probable" that all attempts to do so will have the same defect and fail for the same

reason. In an earlier version of what seems to be the same argument, Frege thus concludes: “Truth seems to be so elementary and simple that a reduction to something even simpler is not possible” (“Wahrheit ist offenbar etwas so Ursprüngliches und Einfaches, daß eine Zurückführung auf noch Einfacheres nicht möglich ist” (NS, 140).)

4

We can try to localise the cause of this epistemic sterility more precisely. The reason why the sense of “true” cannot be eliminated from the concepts revealed by the analysis is that Frege takes it to be constitutive of all acts of grasping a thought, i. e. what he calls “thinking”. The sense of “true” is part of the content of every act of thinking. This is precisely what is expressed by (I). As Frege puts it: “... we can not recognize an object as having a property without, at the same time, acknowledge as true the thought that this object has that property” (*Ged*, 61). (I) only generalizes this point: We cannot even *think* that something has a property without thinking that it is true that it has it.

We can go further and take a closer look at the internal structure of acts of thinking. Thinking, says Frege, is grasping a thought. But if grasping the thought that x is F is the same act as grasping the thought that it is true that x is F , they must have the same content. Thus, (I) implies the following principle concerning contents:

(C) The thought that x is F = the thought that it is true that x is F

All thoughts contain the sense of the word “true” as a part, irrespective of whether they are actually true, false or lack a truth value. Frege is often taken to defend the redundancy theory of truth. But he never argues that the notion of truth can be eliminated in all contexts. On the contrary, as

(I) and (C) make clear, he held the sense of “true” to be omnipresent in our thinking and thus impossible to eliminate.

If we go still a bit further and inspect the internal structure of thoughts, we can see that the sense of “true” is linked to its predicative part. As Frege puts it: “Every property of something is linked to a property of a thought, namely that of truth” (“... *mit jeder Eigenschaft eines Dinges [ist] eine Eigenschaft eines Gedankens verknüpft, nämlich die der Wahrheit*” (*Ged*, 61)). So not only the thinking of complete thoughts but even just thinking of a property of an object (i. e. a concept) necessarily involves the sense of “true”:

(P) To think of a concept *F* implies a grasp of the sense of “true”

We might thus say that, in addition to their specific sense, all concept words express, as their invariable companion and invisible shadow, the sense of the word “true”. Since it is characteristic of analytical definitions to specify concepts (“marks”) in the definiens, the sense of “true” will always be there.

5

Why does Frege believe in the principles (I), (C) and (P)? I think it may have to do with one of his earliest and most stable convictions: that to judge anything is to acknowledge it as true. From this explanation of what judging consists in it is a small step to conclude that the thought that *p* must be the same as the thought that it is true that *p* – which is (C). After all, to acknowledge-as-true that *p* seems to be the same as to acknowledge *that* it is true that *p*, and in this last formulation, “true” is not part of the characterization of the act, but stands to the right of the

that-clause that specifies its content. But if these acts are identical so must be their contents.

The source of the principles (I), (C) and (P) thus seems to be Frege's conception of judgment as the acknowledgement of a thought as true. He suggests that his analysis of judging captures the ordinary sense of the verb "to judge" (or rather: "*urteilen*") as normal speakers understand it (cf. *Vern*, Fn. 4). But this is not only questionable, it has been questioned by many. Bolzano did not believe it and Russell, too argued that the connection between truth and judgment cannot be as tight as Frege says it is. For what seems to be the same reason, Husserl tried to drive a wedge between judging and acknowledging as true. Although Husserl directs his criticism against the neo-kantian philosopher Julius Bergmann, he might very well have had Frege in mind as well (cf. Stepanians, 1998, chap. 12). Indeed, nobody in the Brentano School seems to have agreed with Frege on this point. So it looks as if, at least in this respect, he stood pretty much alone amongst the philosophers closest to him. Of course, to deny against Frege that the sense of "true" is constitutive of thinking is not to deny that there is a close link between the thought that *p* and the thought that it is true that *p*. All the philosophers I just mentioned believed that they are logically equivalent. But they would nonetheless insist against Frege that what we have here are *two* thoughts and not just one.

It seems to me that the interpretation of Frege's argument here presented has some advantages. First, it is not only compatible with (I), but makes it the decisive premise. Second, it takes into account Frege's repeated hints that truth definitions are somehow circular. Third, it explains in what sense the problem arises in the *application* of the definition. And finally, by not presenting Frege's reasoning in the form of a valid argument, it takes seriously Frege's modest claim to have only shown that it is "probable" that truth is indefinable. Perhaps one source

of this modesty is that Frege's argument does not have the required scope. It works only against truth definitions of a certain form, viz. the form of analytical definitions that explain concepts by specifying marks. Frege aims to show that *all* truth definitions fail, but his argument works only against definitions of a quite specific kind.⁴

REFERENCES

DUMMETT, M. *Frege: The Philosophy of Language (FPL)*. London: Duckworth, 1973.

FREGE, G. *Posthumous Writings (PW)*. Edited by H. Hermes, F. Kambartel and F. Kaulbach. Oxford: Basil Blackwell, 1979.

———. *Logische Untersuchungen*. Edited by G. Patzig. Göttingen: Vandenhoeck & Ruprecht, 1976. It contains: *Der Gedanke (Ged)* and *Die Verneinung (Vern)*.

PARDEY, U. *Freges Kritik an der Korrespondenztheorie der Wahrheit. Eine Verteidigung von Freges Kritik gegen die Einwände von Dummett, Künne, Soames und Stuhlmann-Laeisz*. Paderborn: Mentis, 2004.

STEPANIANS, M. *Frege und Husserl über Urteilen und Denken*. Schöningh: Mentis, 1998.

———. *Gottlob Frege zur Einführung*. Hamburg: Junius, 2001.

⁴ This scope-objection has been raised by Wolfgang Künne, *Conceptions of Truth*, Oxford: Oxford University Press, 2003, chap. 3.3.2.