Doubting the Truth of Hume's Principle

Dušan Dožudić

Abstract: Hume's Principle (HP) states that for any two (sortal) concepts, *F* and *G*, the number of *F*s is identical to the number of *G*s iff the *F*s are one-one correlated with the *G*s. Backed by second-order logic HP is supposed to be the starting point for the neo-logicist program of the foundations of arithmetic. The principle brings a number of formal and philosophical controversies. In this paper I discuss some arguments against it brought out by Trobok, as well as by Potter and Smiley, designed to undermine a claim that HP and its instances (such as "the number of the forks on the table is identical to the number of the knives on the table iff the forks are one-one correlated with the knives") are true. Their criticism starts from distinguishing the objective truth from a weak or stipulative one, and focusing on fictional identities such as "Hamlet = Hamlet" or "Jekyll = Hyde." They argue that numerical identities (as occur in instances of HP) are much the same as fictional identities; that we can attribute them only a weak or stipulative truth; and, consequently, that neo-logicists are not entitled to ontological conclusions concerning numbers they derive from HP and its instances. As opposed to that, I argue that such a criticism is ill-conceived. The analogy between the numerical and fictional identities is far-fetched. So, relative to such a criticism, HP has more prospects than some authors are prepared to admit.

Keywords: fictional identities, Hume's Principle, logical truth, mathematical platonism, mathematical truth, neo-logicism, numerical identity, numerical singular terms, objective truth, reference, stipulative truth.

1 Attacking the Principle

The basis of neo-logicist program is the claim that the single non-logical axiom, "Hume's Principle," conjoined with second order logic, yields sufficiently strong system for definition of concept of natural number, relation of immediate preceding, and the natural number zero, as well as for the proof that basic properties of numbers (alternatively captured by Dedekind-Peano axioms) hold.[1] Although second order logic is by itself a matter of considerable controversy, the success of neo-logicist enterprise is usually (or primarily) measured by the ability to defend or discredit Hume's Principle (HP). The principle states that, given two (sortal) concepts, *F* and *G*, the number of objects falling under *F* is identical to the number of objects falling under *G* if and only if the *F*s are in the one-one correlation with the *G*s.

Formal aspect of neo-logicist program is accompanied by ontological, epistemological, and semantic theses intimately related to HP. Numbers are self-subsistent objects, recognised as such and referred to via numerical singular terms.[2] The way we come to know arithmetical truths is via logic, since arithmetic is nothing but logic (except HP, the non-logical axiom). Thus, to decide is a particular instance of identity statement 'the number of *F*s is identical to the number of *G*s' true, we need not appeal to a direct insight into arithmetical realm or some other quasi-perceptual mechanism, but comprehend the one-one relation between the *F*s and the *G*s (explicitly defined in second-order logic). Once we come to know that this correlation holds, and once we know HP – the assumption is that we know it a priori – we can without further ado say that the numerical identity is true, that the embodied numerical singular terms refer, and that numbers as referents of the terms exist.[3]

1 The idea was first presented by Frege (1980). Throughout the book Frege abandoned Hume's Principle as the appropriate definition and introduced it instead as a theorem inferred from the explicit definition of the concept of natural number. The revival of Frege's earlier ideas was first suggested and argued for in Wright (1983). For a formal discussion see Boolos (1998).

2 For example 'the number of objects falling under the concept book,' for short 'the number of books;' or still shorter, '*n*(*BOOKS*).'

3 For a survey and discussion of philosophical aspects of neo-logicism and HP see Demopoulos (1998).

Ever since it was introduced by Frege in 1884, the principle was faced with problems and objections, the Caesar Problem being perhaps the best known example. In the following paper I will focus on a criticism of HP and neo-logicist program based on the claim that HP is too weak for intended purpose. This amongst other things should mean that it does not imply the objective truth of numerical identities contained in its instances, and hence does not imply the existence of numbers as self-subsistent objects.[4] Although doubting the truth of HP motivated by the criticism can be tempting at first, I will argue that such a doubt is ill-conceived and that HP, together with the accompanying philosophical theory, has more prospects as the ground for platonist conception of arithmetic than the criticism suggests. In the rest of this section I survey the mentioned criticism of HP. In two following sections I examine it more closely; I stress its problems and some possible modifications.

Realists praise the objective truth, truth independent of us, our language, intentions and conventions, truth dependent on features of the world (scientific discourse would be paradigmatic here). Some authors further distinguish the objective truth from a weak or stipulative one that depends on our intentions and conventions (fictional discourse is paradigmatic here).[5] The corollary of such a conception of objective truth is that, generally, anyone who accepts it must accept that singular terms embedded in objectively true statements have referents, unlike singular terms embedded in stipulatively true statements (with the obvious exception of negative existentials which are objectively true only when embed empty singular terms). I will call the thesis that singular terms embedded in objectively true statement refer *the referential thesis*. Consequently, since mathematical statements usually are not negative existentials, but do contain singular terms, if

4 Such a criticism is offered by Trobok (2006, 166 – 170). A similar objection can be found in Potter – Smiley (2001, 336 – 338).

5 Trobok (2006, 17 – 21). A similar idea of distinguishing two types of truth can be found in Potter – Smiley (2001, 336). I don't see much point in introducing the two concepts of truth, nor do I regard "stipulative" truth as truth of a substantial kind. Nevertheless, I will retain the terminology in order to avoid additional excurses.

it could be argued that statements containing numerical singular terms are objectively true, it would follow that these singular terms have referents and that mathematical objects exist. But that is precisely an idea behind HP. Isn't, then, HP conjoined with the referential thesis a good case for platonism in mathematics? Some authors believe it is not. For example, Trobok believes that neo-logicists are not entitled to such a conclusion because it is possible

> to maintain that a one-to-one correlation between the knives and the forks does not *itself* suffice for the *truth* of the identity: the number of knives is identical to the number of forks. […] it is an extra step from the truth of this identity to the *objective* truth of this identity in the strong sense, and, hence to the existence of numbers.[6]

Thus, according to the quoted passage, only the objective truth of identity 'the number of knives is identical to the number of forks' (for short '$n(KNIVES) = n(FORKS)$') implies the existence of a number as the referent of the two embedded singular terms. Neo-logicists have thought that applying HP is enough. They were wrong. If '$n(KNIVES) = n(FORKS)$' based on the one-one correlation between the knives and the forks turned out to be true only in a stipulative sense (hence, not *really* true), one would no longer be entitled to conclude that the number belonging to that concepts exists. If that much could be established, it would be a rebutting of neo-logicism. Some authors strive to do precisely that.

But, somebody could wonder, isn't our belief in HP and our general acceptance of its instances an indication that the principle is objectively true? After all, it should be an axiom. What additional non-technical justification could be offered on its behalf? Trobok, together with Potter and Smiley, believes this to be the wrong position. It is true, Trobok admits, that people generally uncritically believe that at least a suitably restricted version of the principle is true, and they are ready to accept on the basis of the one-one correlation between e.g. the knives and the forks on the table that the number of knives is identical to the number of forks. But such a belief need not imply the objective

[6] Trobok (2006, 168).

truth of '*n*(*KNIVES*) = *n*(*FORKS*)' as occurs in an instance of HP. People believe fictional identities such as 'Hamlet = Hamlet' to be true as well, yet scarcely anyone will infer from such a belief that 'Hamlet' refers to Hamlet and that Hamlet exists.[7] She takes this to be an important observation for, equally, no one except neo-logicists would be prepared to conclude that the truth of identity '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP) justifiably implies the existence of a number belonging to these concepts. Why not? Because virtually no one holds that identities such as 'Hamlet = Hamlet,' or '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP), are more than stipulatively true (former identity on the basis of a fictional story, latter on the basis of the stipulation of HP). General reluctance to accept ontological implications brought by HP and its accompanying philosophical theory affirms that.

This conclusion, however, acknowledges a possibility that HP could turned out to be useful after all if an additional criterion, besides our uncritical belief in HP, could be supplied in order to show that '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP) is objectively true (although in that case virtues of HP neo-logicists praise would be considerably reduced). The rest of Trobok's criticism aims precisely at that possibility. Building on her previous conclusion, it is supposed to show that even this can't be done. According to neo-logicists, knowledge that singular terms '*n*(*KNIVES*)' and '*n*(*FORKS*)' refer to the same number, and, consequently, that this number as their referent exists, follows from knowledge that '*n*(*KNIVES*) = *n*(*FORKS*)' is objectively true, and this particular knowledge follow from knowledge that the one-one correlation between the knives and the forks holds. And transition from the one-one correlation to the numerical identity should be justified by HP (*and some other means* if we accept the previously considered argumentation). However, in the light of

[7] "To settle that the statement 'the number of knives is identical to the number of forks' is true, by stipulative reference to a one-to-one correlation between the knives and the forks, is *not* to establish that the numerical singular terms it involves refer. The truth of the identity 'Hamlet = Hamlet' tells us that much" (Trobok 2006, 169).

neo-logicist epistemological assumptions, an additional criterion required leads to circularity:

> […] how can […] identity statements like '0 = 0' be distinguished from identity statements like 'Pegasus = Pegasus'? It seems that, no matter what we insert for '*t*,' the identity statement '*t* = *t*' is always true. One answer might be that 'Pegasus = Pegasus' is not objectively true because 'Pegasus' is not referential. But the point is that we are supposed to know that 'Pegasus = Pegasus' is not objectively true prior to our knowledge the 'Pegasus' is not referential. We are supposed to distinguished identity '*t* = *t*' in which '*t*' is referential from those in which '*t*' is not referential *before* we acknowledge if the term '*t*' is referential or not. This problem, I would say, seems insoluble.[8]

The moral of the passage is that, just as one must know that '*a*' refers in order to know that '*a* = *a*' is objectively true, one must know that '*n*(*KNIVES*)' and '*n*(*FORKS*)' refer, and that they refer to the same object, *before* she comes to know that '*n*(*KNIVES*) = *n*(*FORKS*)' is true. And that is the reverse epistemic order to the one neo-logicists assume when they appeal to HP. The purpose of the appeal to HP was precisely to ensure one that singular terms embedded in its instances refer. To conclude this, however, '*n*(*KNIVES*) = *n*(*FORKS*)' had to be objectively true and one had to know that it is objectively true. But in order to know that the numerical identity is objectively true, according to the quoted passage, one has to know whether '*n*(*KNIVES*)' and '*n*(*FORKS*)' refer, and whether they refer to the same object. However, if one knows that these singular terms refer, and to what they refer, she, contrary to the neo-logicist assumption, doesn't have to invoke HP to find it out, etc. Anyone who concedes such a criticism has no choice but to reject HP as the basis for platonism in arithmetic and neo-logicist program in particular. In the following sections I will examine whether the criticism is justified.

8 Trobok (2006, 170).

2 The attack gone wrong, and its modifications

Comparison of numerical identities as they occur in instances of HP with fictional identities of the form '$x = x$' in the previous section is not a coincidence. To argue that general uncritical belief in the truth of '$n(KNIVES) = n(FORKS)$' based on the one-one correlation between the knives and the forks does not imply the objective truth of that statement, and consequently that the embedded singular terms do not refer, a compelling example of another uncritical belief in the truth of a statement had to be offered. The example had to be similar enough to beliefs that numerical identities (as occur in instances of HP) are true, but at the same time should be obvious that it does not imply the objective truth of a given statement, and reference of its embedded singular terms. If fictional identities of the form '$x = x$' function as previously suggested, believing in their truth would be a good candidate. Namely, if people accept any piece of fictional discourse as *true*, fictional identities such as 'Hamlet = Hamlet' are the one. And since singular terms in such statements obviously do not refer, such statements, according to the referential thesis, can not be objectively true. So this by analogy demonstrates unreliability of general uncritical beliefs mentioned in connection to HP and its instances. The fact that there is a general uncritical belief in truth of numerical identities as occur in instances of HP can not imply the objective truth of such identities.

However, the conception of identity statements of the form '$x = x$' suggested in the previous section as a support to the argument is controversial. As I see it, neither are fictional identities of the form '$x = x$' stipulatively true, nor the objective truth of identities of the form depends on reference of embedded singular terms. Let me elaborate this. Suppose that previously sketched conception of fictional identities of the form '$x = x$' is correct. It would mean that there is (at least for a mathematical platonist) a significant difference between, e.g., '0 = 0' and 'Pegasus = Pegasus.' (The difference would be that the former identity is an objective truth embedding referring singular term '0', and the latter a stipulative truth embedding fictional singular term devoid of reference.) It follows that, since 'Pegasus = Pegasus' is arbitrary stipulation just as 'Jekyll = Hyde' or 'Pegasus = the flying horse' are, it could have happened that Pegasus ≠ Pegasus if somebody stipulated this instead, just as it could have happened that Jekyll ≠ Hyde, or that Pegasus ≠ the flying horse.[9] But there is a difference between fictional identities such as 'Pegasus = Pegasus' on the one hand, and 'Jekyll = Hyde' or 'Pegasus = the flying horse' on the other. While it could have happened that 'Jekyll = Hyde' is not in any sense true (just as 'Jekyll ≠ Hyde' is not, at least in non-modal sense), if 'Pegasus' in both its occurrences in 'Pegasus = Pegasus' is used in the same sense, that identity must be true in the same way '0 = 0' and 'Cicero = Cicero' are. To argue for anything else would be a violation of one of the basic linguistic principles that at least in the same context instances of an expression have fixed meaning (if not indicated otherwise, of course). That principle precedes any stipulation, and virtually any meaningful communication. Otherwise, a language could not function. But if it can not be that Pegasus ≠ Pegasus, while it could have been that Jekyll ≠ Hyde or that Pegasus ≠ the flying horse, 'Pegasus = Pegasus' can not be a mere stipulation. For that reason, it is either a meaningless gibberish or objectively true.[10]

In the light of these remarks some modifications of the initial criticism of HP suggest itself. For instance, a better strategy would be not to claim that some instances of '$x = x$' are not objectively true, but to accept that identities such as 'Hamlet = Hamlet' are objectively true and that some objectively true statements embedding singular terms, besides negative existentials, do not depend on reference of these

[9] A support to the claim that it could have happened that Jekyll is not identical to Hyde is offered in the second part of this section.

[10] As far as I can think of, a conception of fictional identities of the form '$x = x$' according to which such identities are not (objectively) true because occurrences of singular term in them do not refer, closest to that of Trobok, was suggested by Russell (1919, 175 – 176). But, for Russell, statements such as 'Pegasus = Pegasus' come down to quantified statements of the form '$\exists x$ (… $x = x$).' So, the reason why such statements are not (objectively) true is evident: There is no x such that it satisfies particular conditions, not because (apparent) singular term in question does not refer. On the other hand, in her book Trobok makes no commitments to such a view. Hence, we are left without further justification of the position behind her claims about such identity statements.

terms. The previous formulation of the referential thesis, then, should be made more precise. In other words, the initial list of statements containing singular terms to which the thesis does not apply should be extended. With thus modified referential thesis one would be in a position to introduce at least a possibility that '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP) is an objective truth belonging to the list of statements to which the referential thesis does not apply. So, although in that case instances of HP would be objectively true, that would not imply that the embedded singular terms refer, and that numbers as their referents exist.

To justify such an approach we have to make clear what reasons do we have for treating instances of HP as truths of the kind? Certainly, such a possibility can be introduced, but without further support it makes no threat to neo-logicist platonism. The problem is, there are reasons to doubt this possibility, especially if it is merely justified by analogy with instances of '$x = x$' involving fictional singular terms. For one thing, the form of '*n*(*KNIVES*) = *n*(*FORKS*)' is '$x = y$,' not '$x = x$,' and the objective truth of instances of '$x = y$' does in part depend on the fact that the embedded singular terms refer, and that they refer to the same object. So, it seems that the referential thesis should apply to instances of HP after all.[11] Secondly, in case someone would, nevertheless, maintain that '*n*(*KNIVES*) = *n*(*FORKS*)' is not true in virtue of reference of '*n*(*KNIVES*)' and '*n*(*FORKS*)' to the same object, it is not clear in virtue of what would it be true instead. The only alternatives, it seems, would be either to say that it is true in virtue of a stipulation (and in part my paper aims to show that is not the case), or to say that it is true in virtue of the one-one correlation between the knives and the forks. However, this would be a remarkable exception among identity statements of the form '$x = y$,' and it does not seems likely at all. Finally, even if the first two remarks were disregarded, the modified objection appears to be the problem for any platonist conception that embraces numbers as self-subsistent objects, because it indicates the need for a clearer answer to the question to which statements con-

11 I mention a possible counterexample to this remark in the second part of the following section.

taining singular terms the referential thesis applies, and to which does not. I will return to this in the following section.

Since identities of the form '$x = x$' can not be stipulatively true, if one prefers to follow a line of argument set forth in the previous section, she must argue that '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP) is not similar to fictional identities of the form '$x = x$,' but to those of the form '$x = y$' (such as 'Jekyll = Hyde'). That is the suggestion of Potter and Smiley.[12] They claim that identities such as 'Jekyll = Hyde' are not objectively true because neither one of the embedded singular terms refer. But since there is a general agreement that Jekyll and Hyde *are* identical, the agreement grounded in a stipulation, it seems appropriate to classify such statements as "vacuously" true. Numerical identity '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP) should be treated in a similar way.[13] Such a proposal avoids problems of previously discussed argumentation which relied on the specific conception of identities of the form '$x = x$.' But although it sounds tempting to compare numerical identities such as '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP) with fictional identities such as 'Jekyll = Hyde,' differences between the two are too great to be ignored. Upon further reflection it becomes obvious that '*n*(*KNIVES*) = *n*(*FORKS*)' can not be treated as a fictional identity of the form '$x = y$.' Consequently, those two sorts of '$x = y$'-identities can not be true in the same sense. I will point out a few reasons for that.

Let me start with the notion of obviousness. Trobok considers it to be an important argument for the truth-value realism in mathematics.[14] The argument is simple and it goes as follows: At least elementary mathematical statements are obvious and people have strong beliefs whether they are true or false. People's beliefs about truth-values of such statements are in the main correct.[15] So mathematical statements people find true are in the main true, and those they find false

12 Potter – Smiley (2001, 336).

13 Potter – Smiley (2001, 336).

14 Trobok (2006, 21 – 24).

15 That is the additional assumption of the strong truth-value realism that Trobok embraces. See Trobok (2006, 17).

are in the main false. And since people typically find mathematical statements to be true, mathematics *is* objectively true.

The argument is problematic, but for the sake of discussion I will take it for granted. In that case obviousness can be used as a support for HP as well. After all, isn't person's uncritical belief in HP based on the fact that it is obvious to her? Perhaps, but someone could doubt that, bringing our attention to general uncritical beliefs in truths of fiction. However, any relevant similarity between HP and its instances on the one hand, and statements of fictional discourse on the other, disappears as soon as the former class of statements is compared with the later with respect to a degree of their obviousness. Namely, even if it is true in a stipulative sense that Hamlet killed Claudius, it is not obvious. I am told that he killed him, and every correct belief about it is grounded in a common social convention. Tracking down that convention's origins would reveal that Shakespeare did not find it obvious as well, he merely stipulated it. At the same time, I do not believe in, or accept, HP and its instances because somebody told me that the number of *F*s is identical to the number of *G*s if and only if the *F*s are one-one correlated with the *G*s, just as I do not believe that 2+2 is 4 because I have learned in a school that it is. Even if I have at first learned HP or some of its instances, or that 2+2 is 4, by heart, after a while I begin to believe in their truth because I begin to *understand* them, and in process of that understanding I find them obvious and true. If someone had taught me that 1+1 is 3, or that the number of *F*s is identical to the number of *G*s if and only if the *F*s are *not* one-one correlated with the *G*s, I could believe that for a while. But as soon as I would understand what it means, I would discard it as obviously false.

Further reason for not treating HP and its instances as semantic par of fictional statements is this: Granted that people ordinarily accept fictional identities like 'Jekyll = Hyde' as vacuously or stipulatively true, the reason why they accept them as such is grounded in a social convention leading back to the point when those identities were introduced as a part of the fictional discourse for the first time. On the other hand, the reason for accepting HP and its instances as true is not grounded in social conventions, but in the person's *understanding* of the principle. As soon as one understand it, she has to see that the equivalence of the left and the right side is not a stipulation, as is the case with 'Jekyll = Hyde,' but that HP with its instances is true independently of our prior intentions and conventions. Once a person critically reflects on HP and its instances on the one hand, and fictional identities on the other, she *must* see semantic, as well as metaphysical differences.

Finally, this "must" brings me to another important disparity between '$n(KNIVES) = n(FORKS)$' (as occurs in an instance of HP), and 'Jekyll = Hyde.' If both fictional identities and such numerical identities were stipulatively true, placed into modal setting they would induce more or less the same modal intuitions. Yet, it seems that is not the case. I guess that our modal intuitions about identity 'Jekyll = Hyde' suggest something as the following. At the same time Frege was contemplating about variety and validity of abstraction principles, Robert L. Stevenson decided to write a novel about a scientist he named 'Dr. Jekyll.' Dr. Jekyll, Stevenson decided, should have the evil twin called 'Mr. Hyde.' Upon further reflection about human nature (or whatever) and the plot of the story, Stevenson decided it would be more profound and intriguing if Mr. Hyde were not Dr. Jekyll's evil twin, but Dr. Jekyll himself, born in the course of a crazy chemical experiment he conducted. At that point Dr. Jekyll became Mr. Hyde and identity 'Jekyll = Hyde' was created. But Dr. Jekyll could turn out not to be identical with Mr. Hyde if Stevenson had decided to keep Mr. Hyde as Dr. Jekyll's evil twin after all. Or he could have decided to erase Mr. Hyde from the story altogether, and dedicate the novel instead to Dr. Jekyll's Christian virtues. In any case, it *could* have turned out that Dr. Jekyll is not Mr. Hyde and that 'Jekyll = Hyde' does not hold. Since Stevenson decided to stipulate that Dr. Jekyll is Mr. Hyde and publish that version of the novel, today everyone who knows the story believes that Dr. Jekyll is Mr. Hyde and that identity of them holds (or is in some sense true).[16]

[16] It should be noted that this intuition about fictional identities does not contradict results about rigidity, and necessity of identity. Fictional names are not merely descriptive names (in the sense that their reference is secured by a description or descriptions). Descriptions attached to such names are *all* there is. There's no reference, essence, or necessity hidden

Does it make sense to apply the same modal analysis to HP and its instances, and expect a similar result? Do my modal intuitions permit me to construe a consistent story about, say, Frege, in which HP with its instances would turn out to be false? And is it possible to imagine a point in history at which HP with its instances became stipulatively true? And a point before that in which the principle did not hold? I do not see a possible scenario in which any of that could turn out to be the case. And those intuitions seem more compelling than the claim that HP with its instances is merely stipulatively true just as fictional identities are.

So, in summary, up to this point I have stressed some problematic assumptions of the criticism of HP governed by doubts concerning the objective truth of HP and its instances. Further, I have tested potential modifications of such a criticism. That turned out to be unsatisfactory as well. Finally, in the course of my discussion, some arguments in favour of the objective truth of HP and its instances emerged. So, it should be concluded that, given the above discussion, HP is in a better position than the considered criticism suggests. However, still further modifications of the criticism impose themselves. Namely, it can be reformulated without reliance on assumptions that have so far turned out to be problematic. So, my next task is to investigate whether such a modified criticism is immune to remarks of this section, and whether it makes more profound negative effect on HP.

3 Further modifications

The conclusion concerning epistemic circularity of instances of HP mentioned at the end of the opening section relies on the prior conclusion that the one-one correlation itself does not suffice to establish the objective truth of numerical identities such as '*n*(*KNIVES*) = n(*FORKS*).' If that conclusion could be showed to be unjustified, the circularity argument would fail as well. For that reason I will focus on potential modifications of the former argument and see what this amounts to. I will start by explicitly stating it:

behind them. The matter is, though, highly controversial and exceeds the scope of this paper, so I will not pursue it further.

(A1) If a statement containing singular terms is objectively true, embedded singular terms refer to existing things (the referential thesis).

(A2) General acceptance of HP and its instances as true was one of the key reasons for neo-logicists to argue that the principle via its instances implies the existence of numbers.

(A3) In so doing, they have interpreted 'true' as *objectively* true.

(A4) However, there is a set of statements containing singular terms generally accepted as true, but it is not accepted that those singular terms refer (e.g. 'Hamlet = Hamlet'), although at the same time the referential thesis is embraced.

(A5) So, such statements are not objectively, but *stipulatively* true.

(A6) Since it is believed that, e.g., '*n*(*KNIVES*) = *n*(*FORKS*)' (as occurs in an instance of HP) is true, but at the same time there is a general reluctance to accept this to imply that the embedded singular terms refer, '*n*(*KNIVES*) = *n*(*FORKS*)' is similar to statements such as 'Hamlet = Hamlet,' and should be treated as stipulatively true as well.

(A7) If merely stipulatively true, HP is too weak for neo-logicist purposes. The objective truth of 'the knives are in the one-one correlation with the forks' is not enough to secure the objective truth of '*n*(*KNIVES*) = *n*(*FORKS*).' In order to establish whether '*n*(*KNIVES*) = *n*(*FORKS*)' is objectively true, and whether '*n*(*KNIVES*)' and '*n*(*FORKS*)' refer, an additional criterion is required. It is at best an extra step from stipulative truth of HP (and its instances) to its objective truth.

In the previous section I gave some reasons for discarding (A5) and (A6). Still, one could try to save the argument's underlying idea by avoiding commitments to (A5) and (A6). A way to do it would be to insist that, although numerical identities (as occur in instances of HP) can not be treated as fictional identities in general, maybe the reluctance to accept such identities as objectively true, after learning about accompanying ontological commitments, shows *by itself*, given the referential thesis, that HP is not objectively true. So, the new argument against HP would be:

(B1) People generally uncritically believe that HP and numerical identities as occur in its instances are true.

(B2) If a statement containing singular terms is objectively true, embedded singular terms refer to existing things (the referential thesis).

(B3) Generally, no one besides neo-logicists believes that instances of HP imply the existence of numbers.

(B4) Hence, as soon as one is told about commitments of believing in truth of the principle and its instances, she is no longer willing to accept them as objectively true.

The reason why a person can't get rid of the feeling that HP and its instances are, in spite of this, true is that there is a weak concept of truth as well. Just as one believes that 'Hamlet killed Claudius' is true (and does not believe that 'Hamlet' refers to Hamlet, and 'Claudius' to Claudius), so one believes that, say, '$n(KNIVES) = n(FORKS)$' as occurs in an instance of HP is true (although '$n(KNIVES)$' and '$n(FORKS)$' do not refer to a number).

But consider the transition from (B3) to (B4) more closely. The argument read into that transition is this:

(C1) If HP and its instances are objectively true, numbers exist.

(C2) Numbers do not exist (at lest not on the ground of numerical identities embedding numerical singular terms, which occur in instances of HP, that one uncritically accepts as true).

(C3) Hence, HP and its instances are not objectively true.

However, (C)-argument contains a hidden assumption, namely, the premise (B2). Once explicated, (C3) is no longer the only possible conclusion. The fully explicated argument goes:

(D1) If HP and its instances are objectively true *and* if the objective truth implies the referential thesis for statements containing singular terms, numbers exist.

(D2) Numbers do not exist (at least not on the ground of numerical identities embedding numerical singular terms, which occur in instances of HP that one uncritically accepts as true).

(D3) Hence, HP and its instances are not objectively true.

Now, with this explication it becomes apparent that transition from (B3) to (B4) is not the only one possible. Given (D1) and (D2), (D3) is not the only possible conclusion. Instead, by accepting (D1) and (D2), it could be concluded:

(D3*) The objective truth does not imply the referential thesis for statements containing singular terms (or at least a restriction of the domain of statements to which the thesis applies is required).

Or it could be concluded:

(D3**) HP and its instances are not objectively true and the objective truth does not imply the referential thesis for statements containing singular terms (or at least a restriction of the domain of statements to which the thesis applies is required).

The reason why Trobok in transition from (B3) to (B4) assumes (C1) – (C3) is that she never doubted the thesis that the objective truth implies the referential thesis (as formulated in the opening section) for statements containing singular terms (at least statements other than negative existentials).[17] Her application of that thesis to identity statements of the form '$x = x$' clearly demonstrates that. Accordingly, to doubt the objective truth of HP and its instances was the only option left. But neither it is obvious nor trivial that the objective truth implies the referential thesis (at least in its unrestricted form), nor it is obvious that everyone engaged in "(B3) to (B4)"-scenario, thus faced

[17] Negative existential statements are the only statements containing singular terms that escape the referential thesis that Trobok explicitly recognises in her book; see Trobok (2006, 36, f. 38).

with (D1) and (D2), would conclude (D3). In order to claim that (D1) – (D3) is the only acceptable argument and, accordingly, that the step from (B3) to (B4) is the correct one, good reasons are required for accepting the thesis that the objective truth implies the referential thesis in its unrestricted form. And that means at least in a form strong enough to carry a platonist conception, which is to say, in a form that will apply to arithmetical statements.

By this point, however, it is obvious that the unrestricted version of the referential thesis will not do. Hence, if its usefulness and credibility is to be retained, the answer to the question to what kinds of statements containing singular terms it will apply, and which statements should be excluded, must be offered. Do negative existential statements and instances of logical truths exhaust the list of excluded cases? What about arithmetical statements such as '2+2 = 4'? How can I be confident that they do not belong to the list as well? Or '$n(KNIVES) = n(FORKS)$' as occurs in an instance of HP for that matter? Anyone who feels that instances of HP belong to the list of excluded cases would have to offer good reasons for that, and I am not convinced that there are such reasons. Perhaps the only possibility would be to argue that its instances are true in virtue of HP *alone*, but that would be highly contra intuitive and in conflict with what I said in the previous section. It would be similar to the claim that 'Cicero = Tully' is true in virtue of the law that everything is self-identical alone. And that would be false or at least incomplete statement.

Of course, the matter is not straightforward. There are statements containing singular terms that are neither negative existential statements nor instances of logical truths, yet it could turn out that the referential thesis does not apply to them. The example I have in mind is a modification of infamous 'Nothing can be red and green all over (at the same time).' Try to formulate a similar statement using singular terms instead of predicates (or, alternatively, use phrases such as 'is identical to Plato' as complex predicates). We get for example 'Nothing can be Plato and Aristotle.' The assumption is of course that Plato is not identical to Aristotle (but the same is with the former case involving properties). Now, is '$\text{Hamlet} \neq \text{Ophelia} \rightarrow \neg\exists x\ (x = \text{Hamlet} \wedge x = \text{Ophelia})$' any less objectively true than '$\text{Plato} \neq \text{Aristotle} \rightarrow \neg\exists x\ (x = \text{Plato} \wedge x = \text{Aristotle})$,' just because 'Hamlet ≠ Ophelia' is a stipulation and 'Plato ≠ Aristotle' is not? It seems that if one is to insist that the assumed fact about identity affects the objectivity of a particular statement of the form '$a \neq b \rightarrow \neg\exists x\ (x = a \wedge x = b)$,' it would mean that such statements are merely empirical, and yet it looks as if they are not. So, it comes as a conjecture that, although such statements involve two different singular terms, their objective truth does not depend primarily on the reference of these singular terms. But I mention this just as a possibility, to show that there could be an exception, and I will not push the matter further.

Back to our former concern; one *can* argue that instances of HP escape the referential thesis. But in that case she must offer good reasons for that. Those who, nevertheless, feel that such instances should *not* be placed on the list of excluded statements have two further options. They can accept them as full-blooded objective truths committed to the referential thesis. Alternatively, they can proceed still under impression of a notion of a stipulative truth and argue that HP and its instances are such truths. However, to accept any other option but the one that HP and its instances are full-blooded objective truths would require considerable amount of additional work. And, given the above discussion, I am not convinced it would take those ready to pursue it anywhere. Just recall the previous discussion concerning fictional identities. Granted that 'Hamlet killed Claudius' or 'Jekyll = Hyde' are true in a stipulative sense, I can uncritically believe that HP with its instances and 'Hamlet killed Claudius' are both true, but once I invoke my modal intuition about HP and its instances, and about fictional statements, it becomes obvious that they are not true in the same sense. Relying on obviousness should not be disregarded as well, since, if it could be turned into an argument for the truth-value realism in mathematics, it could be turned into an argument for the truth of HP and its instances as well. So, all things considered, there is a good case for the thesis that HP and its instances are objectively true.[18]

18 A version of the paper was presented at the conference *Kritika u filozofiji* in Zagreb at Centre for Croatian Studies, October 2007.

University of Zagreb
Centre for Croatian Studies – Department of Philosophy
Borongajska 83d, 10000 Zagreb
Croatia
ddozudic@hrstud.hr

References

BOOLOS, G. (1998): On the Proof of Frege's Theorem. In: Boolos, G.: *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press, 275 – 290.

DEMOPOULOUS, W. (1998): The Philosophical Basis of Our Knowledge of Number. *Noûs* 32, 481 – 503.

FREGE, G. (1980): *The Foundations of Arithmetic* (tr. J. L. Austin). Oxford: Basil Blackwell.

POTTER, M. – SMILEY, T. (2001): Abstraction by Recarving. *Proceedings of Aristotelian Society* 101, 327 – 338.

RUSSELL, B. (1919): *Introduction to Mathematical Philosophy*. London: George Allen and Unwin.

TROBOK, M. (2006): *Platonism in the Philosophy of Mathematics*. Rijeka: University of Rijeka – Faculty of Arts and Sciences.

WRIGHT, C. (1983): *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.