Abstract. In this paper I offer a conceptually tighter, quasi-Fregean solution to the concept horse paradox based on the idea that the unterfallen relation is asymmetrical. The solution is conceptually tighter in the sense that it retains the Fregean principle of separating sharply between concepts and objects, it retains Frege's conclusion that the sentence 'the concept horse is not a concept' is true, but does not violate our intuitions on the matter. The solution is only 'quasi'-Fregean in the sense that it rejects Frege's claims about the ontological import of natural language and his analysis thereof.

Keywords: concept, object, unterfallen, history of analytic philosophy.

I. Preliminaries

In his Foundations of Arithmetic ([FA]), Frege famously articulates three principles guiding him in the inquiry on the nature of numbers:

always to separate sharply the psychological from the logical, the subjective from the objective; never to ask for the meaning of a word in isolation, but only in the context of a proposition; never to lose sight of the distinction between concept and object. (Frege 1960, p. xxii)

It is the third principle and its implications that will preoccupy me here, though the second one will also feature at various points in the argument. One of the problems that the sharp distinction between concept and object engenders is widely known as ‘the
concept horse problem,’ and was first articulated by Benno Kerry, a contemporary of Frege’s. Crudely stated, it amounted, in Frege’s reply to Kerry, *On Concept and Object* ([C&O]), to the seemingly paradoxical assertion that ‘the concept horse is not a concept.’ (Frege 1960, p. 46)

Now, Frege did not seem to believe that this is a serious problem, and he blamed the awkwardness of the expression on linguistic idiosyncrasy. However, I claim that the problem is indeed a problem, that Frege runs the risk of having his theory of language (and indeed his philosophy of mathematics) undermined by ontological incoherence, and that, ultimately, a scrupulous Fregean will have to drop some assumptions leading to the problem. On the other hand, while I will argue that Frege’s response to the problem is unconvincing, he is not mistaken in believing that the proposition ‘the concept horse is not a concept’ is true.

Let us take the conceptual route that led Frege to this infamous position. There are two levels that need to be considered, the linguistic and the ontological. On the first, one must note that all meaningful expressions, for Frege, are names. According to his completed view, a name has both a sense and a reference (cf. *On Sense and Reference*). Now, some names are complete (or saturated), and the others are incomplete (unsaturated). Complete names are such expressions as proper names (e.g. ‘Gotthlob Frege’), sentences (e.g. ‘Snow is white’) and what we would, in our contemporary jargon, call definite descriptions – expressions like ‘the so and so’ (e.g. ‘the capital of France’). Incomplete names are things like predicates (e.g. ‘… is white’), connectives (e.g. ‘and’), or quantifiers (e.g. ‘for all …’).

With regard to ontology, stuff is divided into objects and functions. Objects include, among others, physical things (e.g. cars, atoms, cities), truth-values (for Frege, there are only two, the truth and the false) and, famously, numbers. Functions include, among others, mathematical functions (addition, derivatives and so on) and concepts. Now, concepts are just like mathematical functions, except their codomain consists only of the two-element set of truth-values. That is to say, when applying a concept to an object
(say, the concept of being white to snow), the result is either truth or falsity (in this case truth), whereas when applying a mathematical function to some object or objects (i.e. numbers) the result is generally another number or n-tuple of numbers (Frege 1960, pp. 30-32).

There is a very rigid connection between the linguistic and the ontological levels, to the effect that, for Frege, all complete names refer\(^1\) to objects, and all incomplete names refer to functions. For example, ‘Gottlob Frege’ refers to the author of *Begriffsschrift*, ‘snow is white’ refers to truth, and ‘the capital of France’ refers to Paris. Then, of course, addition refers to that function which takes a pair of numbers into their sum (7 and 5, for example, into 12) and ‘... is white’ refers to that function which takes all and only white physical objects into truth and all other physical objects into falsity.

II. Problems

We are now in a position to look on the issue proper. There is a tension between the form and the content of ‘the concept horse.’ According to what was explained above, we have the following:

A) ‘the concept horse’ refers to an object (by its form as a definite description);
B) ‘the concept horse’ refers to a concept (by its content, which purports to pick out a concept);
C) No concept is an object, and no object is a concept.

Prima facie at least, the position is incoherent, so one of the three will have to be dropped. Now, the third one should fall only as a last resort, since it is one of the principles of Frege’s philosophy, and the first two are merely consequences of the principles. So there are two simple ways of saving Frege’s

---

\(^1\) I mean, here and in the rest of the paper, to use the verb ‘to refer’ in the strict technical sense given by Frege to *bedeuten*. 
theorising from inconsistency. First, denying that the form demands us to understand that it denotes an object. Second, denying that that object is a concept.

Unfortunately, there are also, \textit{prima facie}, fairly obvious arguments against both. Suppose we wish to deny (A). Then not all definite descriptions are to be understood as referring to objects. This entails that we cannot be certain that such expressions really do refer to objects without a further criterion distinguishing between definite description that refer to objects and definite descriptions that do not refer to objects. How are we to offer such a criterion in a rationally warranted way?

We could, of course, stipulate that expressions like ‘the concept X’ (where X is a concept) do not refer to objects, while all the other definite descriptions do refer to objects. The fact that this solution is simply \textit{ad-hoc} is the least of its problems. There are many other definite descriptions whose special status we will have to stipulate. For example, ‘the universal quantifier,’ ‘the predicate P’ (where P is a predicate), and so on for every definite description of standard \textit{and} non-standard logical operators (consider ‘the and-functor’ in first order logic or ‘the box operator’ in modal logic). It does not seem that we will ever be able to stipulate every kind of exception, all the more so since there are infinitely many logical and mathematical operations which should be functions, but which can be picked out by definite descriptions (‘the derivation operation,’ ‘the sine function’ and so on and so forth).

Further, such a criterion lacking, it seems we would not be entitled to to \textit{use} definite descriptions to pick out proper objects from the world. In order for ‘the current president of the United States’ to pick out Barack Obama, we would need a proof that this definite description refers to an object, or else we would need to stipulate that it does. Eventually, we end up with a completely unprincipled way of using definite descriptions. In ordinary conversation, this will pose no problems; but this would be a disastrous result for science and philosophy, which would lose a primary conceptual tool.
Mirroring quandaries result if we take the second option, namely denying that ‘the concept horse’ refers to a concept. For then, what entitles us to claim that ‘the capital of France’ refers to a capital? Or that ‘the number eight’ refers to a number? Or that either of the two refers to an object? Again, the need for a demarcation criterion appears, only this time at the semantic rather than the syntactic level. The consequences are nevertheless the same.

But if the concept horse is not a concept, we have further difficulties when attempting to spell out the technical details of the proposal. Since the issue is semantic, there is a question of determining the truth of propositions including terms like this. Take the example:

(1) ‘the concept F = the concept G’;

What are the necessary and sufficient conditions for the truth of this statement? Clearly, the statement is true if and only if all Fs are eo ipso Gs and all Gs are eo ipso Fs. Therefore, (1) is true if and only if:

(2) (∀x)(Fx ↔ Gx);

But this is the exact truth-condition of:

(3) ‘F=G’;

The problem here is this: (3) is an identity between predicates (i.e. incomplete names), whereas (1) (in case expressions like ‘the concept horse’ refer to objects and are therefore saturated) is an identity between singular terms (i.e. complete names). Now, if both (3) and (1) are both logically equivalent to (2), then they are equivalent to one another and so they will have the same consequences.

Now, recall that in Begriffsschrift, §3, Frege rejected the logical relevance of the Aristotelian analysis of statements on the grounds that sentences with different subjects and predicates are nevertheless logically equivalent, the now-famous examples being
'The Greeks defeated the Persians at Plataea' and 'The Persians were defeated by the Greeks at Plataea' (Frege 1967, p. 12). That argument runs more or less thus:

- **P1)** If a difference in statement analysis does not entail a logical difference, then the analysis is faulty;
- **P2)** Aristotle analyses statements as composed of subjects and predicates;
- **P3)** There are statements with different subjects and different predicates that are nevertheless logically equivalent;
- **C)** Aristotle's analysis of statements is faulty. Corollary: the subject-predicate distinction is logically irrelevant.

Now, we could formulate a parallel (i.e. not strictly analogous) argument, starting from the equivalence of the identity statements (1) and (3) above.

- **P1)** If a difference in statement analysis does not entail a logical difference, then the analysis is faulty;
- **P2')** Frege analyses statements as composed of functions and arguments/singular terms;
- **P3')** There are identity statements between functions that are equivalent to identity statements between arguments/singular terms;
- **C')** Frege's analysis of statements is faulty. Corollary: the function-argument distinction is logically irrelevant.

This conclusion is clearly unacceptable, since the function-argument distinction, and the logic which is founded upon it, is the cornerstone of Fregean philosophy.

And this brings us to the further conclusion that naïve rejection of (A) or of (B) runs into considerable (and possibly insurmountable) difficulties. Notice also that not all of the difficulties can be removed by rejecting (C), even if we were to consider that alternative viable. A more sophisticated solution is needed, and in the following section I shall consider Frege’s own
solution in [C&O], which amounts to the postulation of proxy objects representing concepts.

III. Proxies

Frege denies that ‘the concept horse’ refers to a concept. In [C&O], however, he proposes more than just this negative thesis. His positive account is that said expression refers to ‘a quite special kind of object’ (Frege 1960, p. 50). As per the explanation of the third methodological principle in the [FA], concepts cannot be made into objects without them being altered in some way, and thus ‘the concept horse’ refers to the ‘objectification’ of ‘... is a horse.’ The object in question stands proxy for, or represents, the concept. In what follows, I shall call such objects ‘proxy objects’ or ‘proxies’.

Let us see whether or not this avoids the problems mentioned above for the rejection of (B). First, we had the issue of the functioning of the mechanism of definite descriptions. Frege’s account defuses the issue by pushing back the semantics to the ontological level. ‘The concept horse’ may not refer to a concept, but it refers to an object that represents the concept we initially think the expression ought to refer to. So, whereas the naïve rejection of (B) simply severs the all-important semantic relation between the expression and the concept, Frege’s account simply makes this relation more complex and indirect. But the chain of reference is ultimately preserved, albeit at the cost of a dubious semantic-like relation holding at the ontological level between concepts and a special kind of objects.

The second issue was undermining the logical relevance of the distinction between functions and arguments, on the basis of the equivalence between statements of identity between functions and statements of identity between arguments. Without Frege’s account, the rejection of (B) left us with only one possible interpretation of (1) ‘the concept F = the concept G.’ This lead us to understand its truth-condition as identical to the truth-condition of (3) ‘F=G.’ But under Frege’s understanding, we can understand
(1) as being true in virtue of the semantic-like connection between ‘the concept F’ and ‘F’! Therefore, (1) has no truth-condition independent of (3), or better yet, (3) is the truth-condition of (1) – and (2) is the truth-condition of (3). Another way to express the same idea is to notice that (1) says nothing over and above (3), and so (1) is to be interpreted as another linguistic form of the same thought expressed otherwise as (3). (1) merely looks like an object identity statement just like ‘the concept horse’ merely looks like it refers to a concept.

At first sight, therefore, Frege’s proposal seems to offer a robust way of getting out of the apparent paradox of definite descriptions for concepts. However, the account needs to be further spelled out. The distinction between concept and object is an ontological one, so there is a matter of specifying more precisely the ultimate nature of the two categories. One needs to answer the questions: what is the nature of concepts? What is the nature of objects? And, given that we have a peculiar kind of objects, the proxy ones, who seem to be different from the rest – what is the nature of proxy objects, what makes them different from ordinary objects?

Frege leaves us in the dark about these questions – he merely postulates that there are unbridgeable differences between concepts and objects and between ordinary objects and proxy objects. A full elucidation of Frege’s ontology is beyond the scope of this paper, and has been done before, notably by Wells (1951). But, on the one hand, as it will turn out later, Wells’ reconstruction proceeded along the wrong lines; and on the other hand, he ignored the issue of proxy objects. I shall therefore now focus on this last issue and attempt to locate proxy objects in Frege’s ontological scheme. As it turns out, the question poses significant problems for Frege’s account.

For what are these proxy objects? Wells points out to several kinds of objects: truth-values, ideas, ranges, and so on (Wells 1951, p. 542). If we are charitable to Frege, we can admit that we have a fairly good idea (or at least a good intuitive grasp) of the kind of stuff this is supposed to be. How are proxy objects any different? Frege merely mentions that they represent the respective concepts
they stand proxy for, and while this may look as if it is not very constraining, we cannot accept that just about any object can stand proxy for any concept. For instance, it does not seem conceptually satisfactory to claim that this chair I am sitting on represents the concept horse, or that my cup of coffee represents the concept book.

If 'natural' objects will not do the job, then perhaps we ought to understand proxies as 'artificial'. Suppose each concept has a proxy object attached, and this object does nothing except represent its respective concept, and is nothing over and above an objectual projection of the concept. This is not satisfactory for two reasons. Firstly, it is obviously an ad-hoc solution and does not seem warranted in any reasonable way.

But secondly, and more importantly, this solution brings with it a prodigious ontological promiscuity. We simply postulate the existence of infinitely many objects, each corresponding to a concept and each of whose nature is exhausted by the function of representing that object. Perhaps we can, adopting a radical platonism, admit this, and it may be that there is such an object for concepts like horse, house, etc. However, this proposal loses all plausibility once we realize we can construct an endless string of concepts from any mundane concept.

Take two concepts, first our old concept horse and then the concept proxy object representing the concept horse. The object representing the first is ‘the concept horse,’ and let us name it $C_h$. Now, the object representing the second is ‘the concept proxy object representing the concept horse’ which is yet different, and which we should represent as $\{C_h\}$. We can continue down this path with ‘the concept proxy object representing the concept “proxy object representing the concept horse”’, which is again different and which we should represent as $\{\{C_h\}\}$. Analogously we can get $\{\{\{C_h\}\}\}$, and so on and so forth; we have an uncontrollable proliferation of artificial objects. This does not make for very robust metaphysics.

Proxy objects have to stand in a closer and more natural connection to their respective concepts. What if we took a specific horse, say Bucephalus, as proxy object for the concept horse, or my
cup of coffee as proxy for *cup of coffee*? Again, this does not seem satisfactory. For why should Bucephalus be the proxy for *horse* rather than any of the other horses, and generally, why should a specific member of the extension of a concept be the proxy for that concept rather than any other member of its extension?

If we take this route, we have to eventually concede that a concept’s proxy object is an arbitrary member of its extension. As it so happens, the usual understanding for the semantic value of a variable is exactly this: that it denotes an arbitrary member of its range; and its semantic role, under e.g. Tarski’s understanding of variables, is to denote the range, i.e. the set of values it can take (Fine 2007, p. 10). The option that suggests itself is that a concept’s proxy object is its extension (Wright 1983, pp. 18-19).

In [C&O], Frege seems to be sympathetic to this idea:

> If he [Kerry] thinks (cf. p. 281) that I have identified concept and extension of concept, he is mistaken; I merely expressed my view that in the expression *the number that applies to the concept F is the extension of the concept like-numbered to the concept F* the words ‘extension of the concept’ could be replaced by ‘concept.’ Notice carefully that here the word ‘concept’ is combined with the definite article. Besides, this was only an incidental remark; I did not base anything upon it. (Frege 1960, p. 48)

If indeed expressions like ‘the concept F’ refer to proxy objects, and ‘the extension of the concept …’ can be read as ‘the concept …,’ then proxy objects ought to be extensions. And while this remark is, in Frege’s words, incidental, other ‘incidental’ remarks in [C&O] lend credence to this interpretation: on page 47, we are given the example ‘The concept man is not empty’; and then, on page 49: ‘The concept *square root of 4* is realized.’ Now, quite clearly, the only things susceptible of being empty or realized are sets. The interpretation of proxy objects as extensions, then, seems to be in line with Frege’s thought. Can it answer our problem?
Unfortunately, it cannot. While so far all our bases look covered, extensions bring with them other problems. The first of them affects the semantic links at least in the cases of co-extensive predicates. Consider the famous contingently-true universal statements, ‘Renates are chordates’ and ‘Chordates are renates’ (or the less technical versions, ‘All creatures with kidneys are creatures with hearts’ etc.). Since the extension of renate is identical to the extension of chordate, ‘the concept renate’ will refer to the same object as ‘the concept chordate.’ But which concept does that object represent? We are inclined to think it represents both, but then we are back the problems in the second section. For we want to be able to say truthfully that

(4) ‘The concept renate is not the concept chordate,’

but on this reading (and in this world where all and only renates are chordates) this sentence is at best simply false, at worst a contradiction, and somewhere in between paradoxical. This happens because this forces us, against Frege, to understand concepts extensionally, in virtue of their reference only, and by abstracting from senses. And while co-extensive concepts of this sort may be rare, there are infinitely many concepts of different sorts which are not realized, and are thus co-extensive in virtue of their extensions’ being empty. For example, under this reading we may be forced to acquiesce to

(5) ‘The concept jars of zakuska consumed by me while writing this essay is the concept square circle’

Chance has made it so that I did not eat any zakuska while writing this essay, so the first object in the identity above is the empty set; logic has made it so that there are no square circles, so the second object in the identity above is also the empty set. The empty set is identical to itself, ergo (5) is true. But this is clearly paradoxical and consequently unacceptable.
While these are issues because the consequences we draw from them are counter-intuitive, there is a further problem which does not appeal to our pre-theoretic understanding of semantics. We can, as it were, ‘russellize’ the notion of proxy object. Consider ‘the concept proxy object that does not include itself as an element.’ As in Russell’s paradox, if the object denoted by the expression between inverted commas in this last proposition includes itself as an element, then it does not include itself as an element; and if it does not include itself as an element, then it includes itself as an element.

It would seem that, either way we turn, we are beset by insuperable difficulties. Is, then, Frege’s third principle to be discarded? In the following section I will suggest that this is not so. But a reinterpretation of said principle is in order.

IV. Principles

It has to be admitted that the connection Frege sees between the linguistic, logical and ontological structures is a bit odd. For why should his theory of language have any ontological grip? Why should it turn out that the way language is organized is exactly the way ontology is organized? And why should purely grammatical aspects of language (such as definite articles) capture the logical aspects of thought?

After all, Frege’s first principle was to distinguish between the psychological and the logical, the subjective and the objective – but natural language is definitely not logical and not objective; it is at best a messy intersubjective, contingent, and constantly changing construct. The only comment Frege passes on this matter in [C&O] is to point out that ‘it is here very much to my advantage that there is such good accord between the linguistic distinction and the real one’ (Frege 1960, p. 45). But this sounds like a classical case of rationalization\(^2\). There are languages with no future tense

\(^2\) In the psychological sense of offering a seemingly rational explanation for decision based on feeling or other irrational mechanisms.
(e.g. Japanese or Sicilian) but we would not base a philosophical account of time on this idiosyncrasy, and then declare it a happy coincidence that linguistic usage is so much in line with conceptual reasoning!

Admittedly, language does offer insights into thought, as from a certain level of abstraction the development of one is entwined with the development of the other\(^3\). But this is to be taken as just what it is – an insight and nothing more. The solution I suggest is for ‘the concept horse’ problem makes use of this resource without reading too much into it. I propose that the distinction between concept and object be understood relationally and operationally.

Consider the following sentences:

\[
\begin{align*}
(6) & \text{ Peter is a student in this class.} \\
(7) & \text{ There are 5 students in this class.} \\
(8) & \text{ 5 is prime.} \\
(9) & \text{ Primality is a property of some natural numbers.}
\end{align*}
\]

I think there is little doubt that, \textit{pace} Frege, the natural understanding of these sentences is that ‘John’ is the object of concept \textit{students in this class}, ‘students in this class’ is the object of concept 5, ‘5’ is the object of concept \textit{prime} and ‘prime’ is the object of concept \textit{property of some natural numbers}. Frege opposes this reading for several reasons.

First, he believes it to be an illusion that a concept can be made into an object without altering it. While he does not initially argue for this in [FA], he points to some reasons for this in [C&O]. Apparently, natural language requires various different constructions to indicate the distinction between what is predicated and that which it is predicated about. Notably, the use of ‘is’ is taken by Frege to be an integral part of predicative expressions (which in turn refer to concepts). So, e.g. in (5), what is predicated of ‘John’ is not ‘student in this room,’ but ‘… is a student in this room.’

\[\text{Cf. Davidson 1974.}\]
Here I will charge Frege again of a linguistic *parti pris*. The ulterior developments of Frege's own logic formalize (6)-(9) above as follows (where 'J' stands for 'John,' 'S' for 'student in this class,' 'P' for 'prime' and 'Pr' for 'property of some natural numbers'):

(6') S(J);
(7') 5'(S);
(8') P''(5');
(9') Pr'''(P'');

The copulas are conspicuously absent, and for good reason: they are inconsequential from a logical point of view. The above examples serve to dispel another Fregean claim: ‘Second-level concepts, which concepts fall under, are essentially different from first-level concepts, which objects fall under. The relation of an object to a first-level concept that it falls under is different from the (admittedly similar) relation of a first-level to a second-level concept.’ (Frege 1966, p. 50) As we can see, while there is a difference between different order concepts, a difference between the relations holding between them and their respective objects is not retained and cannot be defended formally.

With these worries put aside, I can explicit my proposal. It is quite precisely what Frege wants to reject here:

one might, like Kerry, regard an object’s falling under a concept as a relation, in which the same thing could occur now as object, now as concept. The words ‘object’ and ‘concept’ would then serve only to indicate the different positions in the relation. This may be done; but anybody who thinks the difficulty is avoided this way is very much mistaken; it is only shifted. For not all the parts of a thought can be complete; at least one must be ‘unsaturated,’ or predicative; otherwise they would not hold together.\ (Frege 1960, p. 54)

---

4 The apostrophes are meant to indicate the order of a concept (5’ is a second-order concept, P” is a third-order concept, etc.) i.e. a syntactic rule determining which formulas are well-formed on the basis of the kinds of argument a function is allowed to take.
But his argument does not seem to touch the proposal, mainly because he seems to assume that if *unterfallen* is a relation, then it is a symmetrical relation. This need not be the case, and indeed it is not. An object's falling under a concept is a polarized relation, and the predicative component of thought is the one that stands at the unsaturated pole.

The asymmetrical nature of the *unterfallen* relation allows us at once to keep the best of both Frege's theories and of our intuitions. For with this understanding, we can preserve Frege's third principle, the truth of 'The concept horse is not a concept,' and the intuition that 'the concept horse' ought to generally denote a concept!

For notice that in 'The concept horse is not a concept,' ‘the concept horse’ is not at the unsaturated pole of the thought expressed therein. On the other hand, the sentence ‘The concept horse is both a concept and an object’ turns out to be false. Similarly, in this last sentence, our troublesome expression is not at the unsaturated pole, so it cannot act like a concept. Then, of course, in ‘Bucephalus falls under the concept horse’ our expression does refer to a concept, for it is placed at the unsaturated pole.

This solution makes us of the corrective capabilities of Frege's second principle: one ought never to ask for the meaning of a word (may we say: expression?) in isolation, but always in the context of a sentence. Why, then, should we ask what a certain expression refers to in isolation? All expressions, including ‘the concept horse,’ can have their references fixed by the context in which they are used, such that we shouldn’t wonder that the same expression refers at one time to an object, and at another to a concept.

Finally, what is understood by ‘concept’ and ‘object’ on this interpretation? The answer is that nothing is understood by the two terms on their own. Rather, we will understand the relation concept-object as offering us an orientation in ontological reasoning. This is similar (though by no means analogous) to Aristotle's hyle-morphe distinction. A molecule is a morphe of the hyle made up of atoms, while an atom is a morphe of the hyle made of protons, electrons and neutrons; and so on and so forth. Like the
hyle-morphe polarization, the concept-object polarization offers us the tools to think about the ontological structure of the world, without committing us to claims about ultimate natures and the like.

Conclusively, this interpretation of Frege’s third principles allows us to keep the best Fregean solutions while avoiding the pitfalls of a too ontologically committed conceptual stance. While diverging from the letter of Frege’s philosophy, this solution seems to me to be perfectly in line with the spirit of Frege’s doctrine.

REFERENCES