Intrinsic Analysis of Antinomies and Self-Reference

Introduction

This essay proposes a reconstruction of the ideas lying behind Finsler's analysis of antinomical situations. We do not think that it is necessary to give a literal account of his arguments: They can be easily followed in his papers. Instead we sought a basis that the various arguments might have in common. This essay is self-contained. It does not depend on the *results* of Finsler's analysis but gives an independent account of antinomical and self-referential situations.

Since Finsler's first philosophical paper [1925], antinomies have been studied in many different ways. Most authors believe that they require us to make some revisions either in our language or in our ways of thinking (cf. for example Quine [1962]). It is generally agreed that there has been no satisfactory solution to the antinomies; those accounts which take place in a formal system seem too restricted to provide a *complete* analysis.

Formal systems are not used in this paper. As Finsler argued so forcefully, formal systems might not possess the flexibility necessary to deal with the antinomies. We do not want to construct an artificial language in order to avoid the antinomies or, as has been tried recently by Barwise/Etchemendy [1987] to incorporate the antinomies into a formal structure: It is ultimately necessary to diagnose the antinomies on their own ground. We should make a self-contained, or *intrinsic analysis* which does not introduce assumptions foreign to the problem but focuses on the realm from which the antinomies arise. In particular, no new conceptions of truth nor theories about the limitation of human thought are needed. We shall use common logic, the traditional distinction between symbols and their meaning, as well as the distinction between conceptual and perceptual facts.

Summary

Antinomies involve perceptually or conceptually distinct objects which become identified during a line of thought, thus producing a contradiction. The intrinsic analysis keeps the distinct objects separate while allowing one to see how they can become identified. The analysis of antinomies calls for distinctions that are present in

self-referential situations in general. They are easily overlooked, however, if no contradiction arises. – Semantic and logical antinomies have a similar pattern, though they operate in different realms. Their emergence gives insight into the formation and ontological status of concepts.

Part I: Antinomies

1. On the distinction between semantic and logical antinomies

The distinction between semantic and logical antinomies goes back to Ramsey [1926]. *Semantic* antinomies involve explicit reference to symbols or sentences, that is, both linguistic expressions and their meanings are present. *Logical* antinomies, however, involve only concepts and conceptual relations from mathematics and logic.

If we formalize the semantics as in model theory, the real distinction between the semantic and logical antinomies is lost. We will not treat these antinomies formally here. In effect, we claim that mathematical logic does not contribute to the *solution* of the antinomies. This does not mean that mathematical logic has not been enormously fructified through the analysis of the antinomies. It is to say, however, that there can be no *comprehensive* account within formal logic alone of how and why the antinomies arise.

An important result of our analysis is, however, that from a certain point of view, the structure of semantic and logical antinomies turn out to be the same. This point of view varies a great deal from mathematical logic. It is grounded in a detailed analysis of self-reference which lies at the heart of the antinomies.

2. Antinomies, contradictions, distinctions

By an *antinomy* we mean an argument which convincingly leads to a contradiction. By a *contradiction* we mean the conjunction of two mutually negated propositions. One proposition asserts that x has the property E and the other proposition asserts that x does not possess the property E :

(C) $(x \text{ is } E)$ and $(x \text{ is non-}E)$.

Surely we can *write* down such contradictions, as for example:

(12 is divisible by 3) and (12 is not divisible by 3);

but there has not been found any actual objects for which there is a property E such that (C) holds. Of course, the number 12 is no such object.

Evidently, to claim, that a house having red and green spots is both red and non-red does not constitute a contradiction in the sense defined above.

An *antinomy*, in its strictest sense, is not an argument that leads just to the conjunction of two mutually negated assertions, but it establishes the *equivalence* of two such opposite assertions. Therefore, an antinomy establishes the conjunction of two converse implications:

$$(x \text{ is } E \rightarrow x \text{ is non-}E) \text{ and } (x \text{ is non-}E \rightarrow x \text{ is } E).$$

In contrast to these notions, *distinctions* arise from observation. In particular, two objects x and x' are called *distinguishable* in case they are distinct and for some property E the following holds:

$$(x \text{ is } E) \text{ and } (x' \text{ is non-}E).$$

For example, circles and polygons are distinguishable objects by virtue of the latter possessing corners.

In some German philosophical literature (cf. for example Hösle [1986], Wandschneider [1993]) the terms "analytischer Widerspruch" and "pragmatischer Widerspruch" are used for contradictions and distinctions respectively. We shall follow this tradition and introduce the terms "analytical contradiction" and "pragmatic contradiction" for contradictions or distinctions respectively. If "contradiction" has no qualifier, then it stands for "analytical contradiction".

3. Derivation of a semantic antinomy

Let us turn now to a specific semantic antinomy. The following antinomy is a slightly modified version of Carnap's [1934] "liar cycle."

(1) a: b is true.
b: a is not true.

The traditional argument goes like this: In case a is true, then b holds. From this it follows that a fails – contradicting the assumption. On the other hand, in case a is false, then b fails. This makes a true – another contradiction. Thus we have an antinomy, the equivalence of two mutually negated assertions. Without doubt, the conclusion of this argument is a contradiction. An intrinsic diagnosis must by its very nature take hold of the course of the argument itself; it must reveal the root cause of the contradiction, not just a way to avoid it.

4. Diagnosis of a semantic antinomy

The argument deriving a contradiction in Carnap's Liar cycle begins with the first line of (1). There is a proposition a, which involves a sentence b in the second line; b in turn mentions a. For the derivation of the contradiction it is necessary that the a in the first line is identified with the a in the second line. *Without this identification there would be no contradiction.* Now, is this identification proper? The only essential property of a present in the second line is that a is the subject of a proposition. In the first line a does not stand for the subject of the proposition, but represents the whole proposition. Thus, the two a's have a different meaning. Hence the two objects, called "a", in the first and the second line of (1) serve a different purpose and need to be distinguished clearly. Since for b we may argue along similar lines, we are presented with a new version of (1) that makes this distinction explicit:

(1') $a^{(1)}$: $b^{(1)}$ is true.
$b^{(2)}$: $a^{(2)}$ is not true.

By strictly following the *principle of identity,* which says that only objects possessing identical properties may be identified, the contradiction evaporates. Only when $a^{(1)} = a^{(2)}$ and $b^{(1)} = b^{(2)}$ are assumed, ignoring the proper distinctions, can a contradiction result.[1]

The common form of the Liar is: "This sentence is not true." Here "this sentence" is the subject of the proposition "This sentence is not true" *and* refers to this proposition as a whole. If we abbreviate "this sentence" by "s" we can represent the full structure of "This sentence is not true" by:

(2) s: s is not true.

---

[1] The essential features of this diagnosis where first pointed out to me by Werner A. Moser.

Here too we have to distinguish between the *s* before the colon and the *s* after the colon. The latter represents the subject of a proposition while the former stands for the proposition itself. Hence it is necessary to indicate the different meanings by different symbols as before:

(2') $s^{(2)}$: $s^{(1)}$ is not true.

Without identifiying $s^{(1)}$ and $s^{(2)}$ no contradiction results.[2]

5. Some objections

One could argue that $s^{(1)}$ designates not only the subject of the proposition $s^{(2)}$ but *also* this proposition itself. Hence $s^{(1)}$ and $s^{(2)}$ are the same. But in this case, the meaning of $s^{(1)}$ (as well as the meaning of $s^{(2)}$) would not be not unique any more. Namely $s^{(1)}$ has two mutually incompatible meanings: On the one hand $s^{(1)}$ is the subject of a proposition and stands on that account for a *part* of the proposition and therefore is distinguished from $s^{(2)}$. On the other hand, $s^{(1)}$ stands for the *whole* proposition and in this role is identical with $s^{(2)}$.

If we were to assume, in order to identify $s^{(1)}$ and $s^{(2)}$, that $s^{(1)}$ represents simultaneously the subject of the proposition *and* the whole proposition, then $s^{(2)}$ must also be given a new meaning. In this case, $s^{(2)}$ is supposed to represent not just a proposition with a subject and a predicate, as before, but also this subject itself. In other words, $s^{(2)}$ is a circular proposition. Obviously, this throws us back into the situation of (2), taking *s* to signify the proposition as well as the subject of this proposition.

A moment's thought shows that this interpretation is incomplete. We would have to take into account the new situation that arises from the identification of $s^{(1)}$ and $s^{(2)}$. Now, the *s* in front of the colon does not only represent the proposition and its subject, as does the *s* that follows the colon, but also a proposition about a proposition. This argument shows that we have to write:

---

2 Goddard/Johnston [1983] argue along somewhat similar lines. They realize that in order to derive the antinomy, one has to assume the identity of two structurally distinct parts. Because they work within predicate calculus, the scope of their analysis is more limited than ours: an intrinsic analysis demands that the nature of the predicates be analyzed; this goes beyond predicate calculus. Hence, Goddard/Johnston cannot explain why and how but only *that* differentiations have to be made in order to derive a contradiction.

(3) $s^{(2)}$: [$s^{(1)}$: $s^{(1)}$ is not true].

But our analysis above shows that this representation is not sufficiently precise, because the two instances of $s^{(1)}$ are distinct and therefore cannot have the same meaning. When we apply the same procedure that was used with (2) to the expression within the brackets of (3), we arrive at

(3') $s^{(3)}$: [$s^{(2)}$: $s^{(1)}$ is not true].

As in (2'), this expression does not give rise to a contradiction, as long as we do not ignore the distinction between the increasingly numerous instances of s.

The transition from (2') to (3) arises from the desire to eliminate any distinction between $s^{(1)}$ and $s^{(2)}$. So we would have to give $s^{(1)}$ an additional property that naturally belongs to $s^{(2)}$ in order to blur the distinction between them. But this effort does not succeed. Since the main structure of (2) is represented within (3), we must substitute (3') for (3). Continuing this process we would have to give $s^{(2)}$ the properties of $s^{(3)}$. In turn this gives a new expression $s^{(4)}$ that has an additional property not possessed by $s^{(3)}$, and so on.

(3_3') $s^{(3)}$: [$s^{(2)}$: $s^{(1)}$ is not true].
(3_4') $s^{(4)}$: [$s^{(3)}$: [$s^{(2)}$: $s^{(1)}$ is not true]].

.
.
.

(3_n') $s^{(n)}$: [$s^{(n-1)}$: [...[$s^{(2)}$: $s^{(1)}$ is not true]...]].

The infinite regress never forces us to the the conclusion that there is not really any distinction between $s^{(k)}$ and $s^{(k+1)}$: for any k = 1, 2, 3, ... The only way that a contradiction can ever appear is for us to assume, contrary to the arguments given above, that

$$s^{(1)} = s^{(2)} = \dots = s^{(n)} = \dots .$$

The contradiction cannot be constructed unless we ignore a factual distinction.

It is often argued that the expression

(2_2) s: s is not true;

lends itself to an iteration if we substitute for s the proposition: s is not true. This substitution gives the sequence that follows.

(2_3) s: [s is not true] is not true;

.

.

.

(2_n) s: [...[[s is not true] is not true]...] is not true.

In order to make the different steps of this iteration more explicit, we write:

(3_2) $s^{(2)}$: $s^{(1)}$ is not true.
(3_3) $s^{(3)}$: [$s^{(1)}$ is not true] is not true.
(3_4) $s^{(4)}$: [[$s^{(1)}$ is not true] is not true] is not true.

.

.

.

(3_n) $s^{(n)}$: [[...[[$s^{(1)}$ is not true] is not true]...] is not true] is not true

In iterating (2_2) it seems at first that we do not need to add any new property to this expression but just carry out what is already present. Hence (2_2) and (2_n) would be equivalent in spite of the fact that s in (2_2) has a simple subject and s in (2_n) has a much more complicated one. Obviously, this leads to a distinction, or "pragmatic contradiction".

We must observe, however, that the equivalence of (2_2) and (2_n) can only be established if we let $s^{(1)} = s^{(2)} = \ldots = s^{(n)}$ in (3_2) to (3_n), for all n. In fact, the different s's do not merely indicate different steps of the iteration but signify, according to our previous analysis, different meanings as well. This has already been shown for (3_2) in section 4.

Now consider the iteration itself. Each step is carried out by substituting $s^{(1)}$ for $s^{(2)}$. This definitely *changes the meanings* of the s in front of the colon with every step. In particular, the transition from (3_2) to (3_3) rests upon the fact that the s in front of the colon no longer signifies a proposition, but a proposition about a proposition.

What looks at first like an identification of $s^{(1)}$ with $s^{(2)}$ turns out to be the cause for a shift in the meanings of the s's in front of the colon. In fact, the different meanings of $s^{(1)}$ and $s^{(2)}$ are the driving forces of the iteration. Only by giving $s^{(1)}$ the meaning of $s^{(2)}$ can the iteration be started and kept going. Therefore, without taking into account these different meanings, there would be no iteration at all. We would be stuck with (2_2) and could never arrive at (2_n). If we deny the different meanings of the s's before and after the colon, then we have no complete argument to actually carry out the iteration. Denying these different meanings and producing the iteration are incompatible: They lead to inconsistent arguments.

The essence of the intrinsic analysis of the Liar antinomy consists in pointing out that distinguishable objects in the sense of section 2 are identified and hence have no identity as something individual. Not observing this difference is an offense against the principle of identity. Therefore, what is real are only the given distinct objects: The *antinomy* is not a fact of reality. The *antinomy* is *created* by the one who derives a contradiction while projecting distinguishable objects into the conceptual realm.

6. Epistemological analysis of a semantic antinomy

In this section it will be shown that a thorough analysis of the argument leading to a contradiction from the Liar cycle (1) includes epistemological categories. It will turn out that the *paradoxer*, namely the one who attempts to derive a contradiction, *actually* does differentiate between the different meanings of a and b in the first and the second line of (1). This differentiation is forced upon him. But later in the argument he chooses to ignore this distinction. The irony of the situation is that the paradoxer observes this distinction even while reasoning as though the two different uses of the parameters are identical. In particular, his observation requires the actual perception of concrete objects.[3]

(1) a: b is true.
b: a is not true.

The assertion, a is true, i. e., [b is true] is true, can be applied to the second line of (1) if and only if we take it in the sense: a is *actually* true. Otherwise it would be an abstract proposition with no consequences for the b in the second line. Hence the a in the first line should be taken as an actual proposition about b: b *is* true. The concrete inspection of b in the second line shows that a is merely the subject of a proposition. The second line, standing alone, does not reveal that a is an *actual* proposition about something. As soon as the a in the first line and the a in the second line are *identified*, the

[3] According to Barwise/Etchemendy [1987] this means that the paradoxer has to take into account the situation the proposition is about, namely itself, taken as a concrete object of the world: a linguistic expression. Situations may include propositions, but are not propositions themselves. They are not purely conceptual but involve some kind of perception.

observed distinction is lost. (The same reasoning applies when a is taken to be not true.)[4]

An analysis which relates the first *and* the second line of the Liar cycle (1) has to take into account the different significance of the two instances of a, likewise for b. This different significance is, in effect, conceptual, but its conception involves a *perceptual* component. Our analysis cannot be purely conceptual in the sense that although the *result* might be purely conceptual, the object of our analysis is not.

The fundamental contrast appearing here is *not* that of two contradictory assertions but rather a pair of distinguishable objects in the sense of section 2. In effect, if the two distinguishable objects are projected into the purely conceptual realm, and if we also ignore the principle of identity, a contradiction does arise. More precisely, the a in the first line of the Liar cycle (1), called $a^{(1)}$, has the property, say E, not only to be a proposition, but to be an actual true proposition *about* the b in the second line. In short: $a^{(1)}$ is E. On the other hand, the a in the second line, called $a^{(2)}$, is merely the subject of a proposition not having a concrete significance. To sum it up: ($a^{(1)}$ is E) and ($a^{(2)}$ is non-E). This conjunction of two propositions has the form of a distinction in the sense of section 2. The identification of $a^{(1)}$ with $a^{(2)}$, namely the identification of the a with concrete signficance, called $a^{(1)}$, with the a with merely conceptual significance, called $a^{(2)}$, leads to a contradiction: (a is E) and (a is non-E).

We conclude that the antinomy, namely the argument leading to a contradiction, starts from a projection of two distinguishable objects (pragmatic contradiction) into the conceptual realm, that is, into the realm of logic. In the next step one neglects the distinction just made and thus ignores the principle of identity. The result is an analytic contradiction.

Let us discuss from this perspective Finsler's [1925] favorite example of a semantic antinomy. Consider a box with the following inscription:

> *1, 2, 3*
>
> *The smallest natural number*
> *which is not specified in this box.*

---

[4] Again, in the terms of the analysis of Barwise/Etchemendy [1987] the above argument amounts to the following: the paradoxer does in fact suppose a particular situation for "a" in the first line in that he applies this proposition to the second line. But later in his argument, the paradoxer acts as if he never made this supposition.

Which is this smallest number? There are only finite many numbers specified in the box. Hence there must be a smallest. Assume it is the number 4. But then 4 *is* specified in the box – hence it cannot be equal to 4. Conversely, assume this number is not equal to 4 but, say, equal to 5. It follows that 5 is *not* the smallest number which is *not* specified in the box. This gives a true antinomy: If the smallest number which is not specified in the box, call it x, is equal to 4, then it is not equal to 4. Conversely, if x is not equal to 4, then it is equal to 4.

Let "T" denote "specified on the blackboard" and let "a" denote "the smallest natural number >3". Then the propositions

$$b\text{: } a \text{ is not-}T,$$
$$a\text{: } b \text{ is } T,$$

have the same structure as the antinomy above. What is relevant for the existence of this smallest number >3 is the fact that it is written in the box, i. e. has the property T. Everything which is not explicitly written in the box is irrelevant. This means that the corresponding proposition is not true. Hence we conclude that this antinomy has the same structure as (1) if we substitute "T" by "true". Conseqently, the same method of analysis applies.

In short, the antinomy is produced by the pragmatic contradiction between the propositional content of "b: a is not T" and the fact that b *is* written in the box. Identifying these two instances of b produces an analytical contradiction.

7. A distinction without antinomical character

The ideas introduced to analyze the Liar are of importance also for assertions without antinomical character but which nevertheless are problematic. Consider the assertion, sometimes called the truth-teller:

(4) $$s\text{: } s \text{ is true.}$$

No one finds a contradiction here: (4) is not an antinomy. And yet here too we must distinguish between the s in front of the colon and that which follows it. The latter, called $s^{(1)}$, is merely the subject of a proposition, and the former, called $s^{(2)}$, stands for a concrete proposition:

(4') $$s^{(2)}\text{: } s^{(1)} \text{ is true.}$$

In this case, the identification of $s^{(1)}$ with $s^{(2)}$ does not lead without further ado to a contradiction, but even so is as unjustified as in the antinomical form (2'). The previous discussion of antinomies requires ideas which are forced upon us by consideration of non-antinomical statements as well. The antinomies serve to introduce these ideas to our attention. It is unnecessary for us to develop strategies for escaping from antinomies. After all, the corresponding contradictions only arise out of our neglect of the above mentioned distinctions. The antinomical result deduced from the Liar situation is *our* creation, not something alien we have to defend against.

Clearly, $s^{(1)}$ and $s^{(2)}$ are distinguishable objects in the sense of section 2. By projecting $s^{(1)}$ and $s^{(2)}$ into the conceptual realm and simultaneously identifying them a contradiction arises, even with the truth-teller (4'). The truth-teller is, strictly speaking, not antinomical, but one still has to deal with distinguishable objects.

Part II: Self-reference

8. Self-referential assertions

The crucial problem presented to us by the antinomies is the *structure of self-reference.*[5]

The common feature of (2) and (4) that makes necessary the transition to (2') and (4') respectively is the self-referential structure of the corresponding assertions. Let E be any property. An *assertion* is called *self-referential,* if it has the following structure:

(S) $\quad s\colon\ s$ is E.

An essential *ingredient* of a self-referential *assertion* is a proposition whose subject is this proposition itself. A proposition standing alone cannot be self-referential.[6] It is essential here that part of the proposition, namely the subject, is associated with something outside the proposition. We are led beyond the proposition itself, understood as a mere conceptual entity, to the

---

[5] This point of view has also been put forward by Kesselring [1984: 104f.].

[6] This is to say that a proposition without any indication to its appropriate situation, or concrete significance, cannot be self-referential. In particular, the situation of self-referential assertions include their own linguistic expression (cf. Barwise/Etchemendy [1987: Chapter 8 and 9]).

situation the proposition is about. The reasoning used to analyze the Liar applies here too; we must pass from (S) to

(S') $s^{(2)}$: $s^{(1)}$ is E.

Doing so, the self-referential structure is apparently lost. This means nothing more than that the thing that refers and the thing that is referred to are not the same. In reality, $s^{(1)}$ and $s^{(2)}$ are not arbitrarily different things, but they are different representations of one and the same thing in reality which comprizes both in a union, building a whole.

Consider the following example:

(5) a: a is an English sentence.

In this case one has to distinguish three things. First, the interior "a" is the subject of a proposition which claims about a that it is an English sentence. Second, the exterior "a" designates this proposition. Third there is the claim that a is, in fact, an English sentence. Hence we are dealing with the question of whether the predicate expressed about a applies to a taken as the whole, that is, whether the conceptual content of the proposition applies to its linguistic representation. The concrete union of the conceptual content of this proposition with its actual linguistic representation is the same union spoken of in (S') above in greater generality.

This concrete union does not arise, if the conceptual content of the proposition does not apply to its linguistic representation, as for example in:

(6) a: a is a Chinese sentence.

However, both expressions (5) and (6) lead to a contradiction as before, if we ignore the different meanings of the interior "a" and the exterior "a". In particular, a symbol having distinct concrete significances is projected into the conceptual realm, leading to a contradiction (cf. section 6 and 7). The contradiction is more obvious in case of (6), since the exterior "a" quite clearly is not a Chinese sentence, but in principle the analysis of (5) and (6) is the same as before.

Note that (6) is merely self-contradictory in the sense that its propositional content does not apply to its linguistic expression: But it is not antinomical, since we cannot derive a logical equivalence of mutually negated propositions.

Consider now the self-referential statement (S) where self-reference itself is denied:

(7) s: s is not self-referential.

The expression (7) arises from (S) if we replace the property E by the negation of self-reference. The contradictory character of (7) is not surprising, because the structure of this assertion is actually self-referential while that is denied by the conceptual content of the proposition.

In (6) and (7) one has to observe *two* overlapping contrasts. First, there is the now familiar contrast between the interior and the exterior "a" or "s" respectively. They are distinguishable objects in the sense of section 2. Another contrast, however makes itself manifest in these examples. When the reader understands the proposition involved in (6) and (7) he notices a conflict between their conceptual content (meaning) and their very form (linguistic expression). The conflict in (6) arises out of the accident of what language is used. One can construct other such conflicts involving accidents of representation, such as:

a: a has fewer than four words.

The conflict appearing in (7) lies in the fact that the conceptual content of the proposition is at variance with the structure of the linguistic expression that carries it.

Note that this contrast between the conceptual content and the linguistic expression of a self-referential assertion also involves two distinguishable objects in the sense of section 2. The projection of *this* contrast into the conceptual realm leads to a contradiction in both cases (6) and (7). This is a common property of all self-referential assertions of the form (S).

According to Barwise/Etchemendy [1987] one may say that the following self-referential assertion,

(8) s: s is self-referential,

signifies its own situation. It is not only a proposition, but a *descriptive proposition* about a situation which it explicitly refers to, namely its own linguistic occurrence. However, this is also true for (6) and (7). The situation of (8) consists only of facts, in particular, of linguistic expressions. In this case, the propositional content, in fact, applies to the situation it is about. Hence, in this sense, (8) is true. As we shall see more clearly in the next section, (6) and (7) are false.

Note that self-referential *assertions* are not propositions in the usual sense but linguistic expressions of a certain structure which relate a propositional content with its own linguistic expression. We may call them *self-descriptive propositions*. In other words, they are propositions which are related uniquely to a specific situation, namely the situation which includes their linguistic expression. (For a more general definition of descriptive propositions in contrast to conceptual propositions, see section 9.)

The relation of self-referential assertions to their own linguistic expression cannot be made explicit in all its implications by any symbolic notation. For example, using the same symbols in (8) for both instances of s before and after the colon is misleading, but expresses the self-referential structure of (8) in a most natural way. Introducing different symbols for these s's seems to destroy just this self-referential structure. The only sensible thing to do is to combine these two notions. This amounts to treating the two instances of s as different (more precisely, distinguishable) but also as equal in the sense that they are both manifestations of one and the same whole.

Ignoring these distinctions does no harm in dealing with self-referential assertions which are not antinomical.[7] We just loose some aspects which might be important to keep in mind if we want to understand the nature of self-reference in a deeper sense.

9. The structure of self-referential assertions

The considerations of the last section make us aware of the basic difference between the labeling part and the propositional part of a self-referential assertion. In particular, we have the linguistic expression and its intended conceptual content. Self-reference is only possible if the thing which refers (the referrer) and the thing which is referred to (the referent) are different. Otherwise there would be no reference at all, only monotonous identity. For instance, in (5), the conceptual content of the proposition and the linguistic expression are both representations of one and the same whole, namely the object under consideration. In effect, what has to be referred to each other are two representations of one and the same thing.

---

[7] There even are consistent formal theories which include self-referential structures which do not take explicitly into account all our distinctions. However, they must in one way or the other cope with the self-referential antinomies. See Barwise/Etchemendy [1987] for a particularly elegant treatment of self-referential assertions and their model theory and Aczel [1988] for the mathematics of self-referential (i.e. non-well-founded) sets. For a characterization of what makes a self-referential assertion antinomical, see Wandschneider [1993, §§ 3 and 4].

These two representations are in (5), and a fortiori in (S), in a special relationship. Namely, the actual sentence has a structure corresponding to the conceptual content of the proposition. It is, so to speak, an instance of the latter.

The structure of self-referential assertions is such that they are not propositions in the ordinary sense, where one has only to deal with conceptual entities; but they relate two representations of one thing, namely the actual instance and its corresponding conceptual structure. Propositions which connect concepts without refering to their instances will be called *conceptual propositions*. For example, the following proposition is conceptual: 12 is divisible by 3. As soon as we distinguish between two different representation of an object, we leave the purely logical realm. Because, logic deals only with *patterns* of representation, that is, *possible* instances and not, as it is the case here, with *actual* representations (or instances).

Hence, self-referential assertions are not part of pure logic: They are not conceptual propositions but so-called *descriptive propositions*. It is no wonder then that purely mathematical accounts of self-reference are wanting in one way or the other.

Descriptive propositions deal with the question of whether an observed real object is an instance of a given concept or not. In the first case one says that the concept applies to this object, or that this object is an instance or an exemplification of this concept. Self-referential assertions are *descriptive* propositions about their own linguistic expression, disguised as *conceptual* propositions about themselves. In order to make this more explicit, we consider the following new notation.

If we denote by "A" a concept and by "a" any of its instances, then we might use an arrowhead, $\triangleright$, for expressing the fact that a is an instance of A in the following way: $A \triangleright a$. This expression signifies the *union* between A and a which is neither identical with the concept A nor its instance a. Consider the following example for this structure: Take K as the concept of the sphere in 3-space. If k is a free flying soap-bubble, then we might write: $K \triangleright k$. Be prepared to differentiate clearly between K and k. The assertion $K \triangleright k$ means that an actual thing, called k, is an instance of the concept K. In this case, since K applies in reality to k, we might speak of the concrete union of K and k, constituting a whole.

Let us return to self-referential assertions. Their basic structure is expressed by

(S) $$s\colon\ s \text{ is } E,$$

where E is any property which can reasonably be applied to a linguistic expression. Our earlier analysis shows that (S) is a self-descriptive proposition. In other words, (S) is a descriptive proposition such that its propositional content applies to its linguistic expression. Using our new arrowhead notation, this can be expressed by

$$E \rhd \text{"}s \text{ is } E\text{"}.$$

Conversely, any descriptive proposition which has this structure can be expressed as a self-referential assertion.

The expression on the left side of $\rhd$ always denotes a conceptual entity (in this case carrying propositional content); the expression on the right side of $\rhd$ denotes a concrete object (in this case a linguistic expression) which is an instance of the conceptual content on the left side.

We can still go one step further in playing with the expression of self-reference. The following expression mentions self-reference and also *is* self-referential according to the definition in section 8. Its structure is an instance of its own propositional content:

(8) s: s is self-referential.

Now, "self-referential" can be replaced by "refers to s", giving

(9) s: s refers to s.

Once again the method of intrinsic analysis leads us to discriminate among the instances of s: Both occurrences of s on the right side of the colon are ingredients of the propositional part, they constitute its subject and predicate. The s preceding the colon designates this proposition itself. Hence we have

(9') s: $s^{(1)}$ refers to $s^{(2)}$.

It should be clear from the intrinsic analysis of (9) that its propositional part, namely

s refers to s

is not a purely conceptual proposition. Because, in this case, the referrer and the referent would need to be conceptually different in order to express a relation and not a monotonous identity. In fact, this proposition is descriptive, namely self-descriptive, and it expresses the fact that a concept s refers to an instance of it, so using the arrowhead notation: $s \rhd s$. In addition, according to (9), the very structure of s expresses self-reference. In other words, s also expresses the concrete union, or whole, of its constituent parts which consists of s taken as a conceptual entity (which is in this case the concept of self-referential assertions) and s as a concrete instance of its propositional content. Hence our symbolic notation produces

(10) $$s\colon\ s \rhd s.$$

Once again, the self-referential structure emerges instantly from (9) and (10), as long as we do not use different symbols for the three instances of s. But our previous analysis shows that the exploration of the fine structure of this self-referential assertion forces differentiations upon us which cannot be ignored unless we forgo, in effect, the very nature of self-reference. From this follows that we cannot stick to (10) but must introduce the differentiation used in (9'), giving

(10') $$s\colon\ s^{(1)} \rhd s^{(2)}.$$

In this expression, $s^{(1)}$ corresponds to the $s^{(1)}$ in (9') and represents the conceptual content of the propositional part of this self-referential structure. The concrete instance entering into this expression is represented by $s^{(2)}$. The fact that $s^{(2)}$ is a concrete instance of $s^{(1)}$ makes s a descriptive proposition, more precisely, a self-descriptive proposition.

The symbols $s^{(1)}$, $s^{(2)}$, s do not denote arbitrarily different things. We introduced these different symbols to aid our analysis. These three symbols refer to different aspects of one and the same whole, i.e., one and the same real object, namely the thing under consideration. In (10') we have a self-referential assertion with the property that its propositional content states exactly its own self-referential structure. Symbolically, s expresses the fact that it is self-referential, namely that $s^{(1)}$ applies to $s^{(2)}$.

It is now easy to represent in our new arrowhead notation a self-referential assertion which is contradictory:

(11) $$s\colon\ s \rhd \text{non-}s.$$

(11) is equivalent to (7) and arises from (10) by negating the fact that s, taken as a self-referential assertion, is an instance of its conceptual content. In effect, the descriptive proposition in (11) denies explicitly its very self-referential structure. Therefore, the conceptual content of this proposition is incompatible with its actual structure expressed linguistically. In other words, (11) is structurally self-descriptive but the concrete instance does not match the conceptual content of the descriptive proposition. Hence, in this sense, (11), taken as a self-descriptive proposition, is false.

This analysis of contradictory self-referential structures is what Finsler [1925] and [1944, §2] might have had in mind, saying, that contradictory assertions of the form (11) are "solvable". In particular, he said that in every such expression the explicit conceptual content contradicts its (conceptually) implicit linguistic structure. However, from our point of view, there is no prima facie contradiction, but distinguishable objects in the sense of section 2 which produce a contradiction only if projected into the conceptual realm.[8]

10. Self-referential concepts

In an intrinsic analysis we are aware of our patterns of thought. We recognize different aspects of self-referential assertions as our mind shifts back and forth between the linguistic expression and its propositional content. This makes us aware that self-referential assertions are disguised as purely conceptual propositions: But we know, in fact, from observing our own thinking, that they are partly descriptive.

Prima facie, a self-referential assertion appears as an unstructured unity. Our analysis finds distinct components in this assertion. Our reflection discovers a complex unity comprising the components which our analysis has identified.

A severe challenge lies waiting to test the method of intrinsic analysis. There are, after all, the logical antinomies which seem to involve, according to section 1, only concepts and conceptual relations from mathematics and logic. The intrinsic analysis of self-referential assertions observed the mind at work, and identified a conceptual component (propositional content) as well as a perceptual component (linguistic expression). In all examples of antinomies discussed so far, the perceptual component is drawn on sense-perception. For instance, the left-hand labels in example (1) are visually associated with the subject of the proposition on the right side. Clearly, perception is at work here.

The logical antinomies, however, challenge us by lacking, prima facie, a perceptual component. They seem to be purely conceptual. But that does not mean, however, that an intrinsic analysis of logical antinomies is confined to purely deductive reasoning.

---

[8] There is a remarkable similarity between Finsler's treatment of the Liar-type antinomies and Buridan's ideas about self-reference (cf. Hughes [1982]). Buridan too does work within the realm of classical (absolute) logic and is convinced that every proposition is either true or false. Concerning the Liar, he comes to the conclusion, that it must be false. His argument rests upon the distinction between a sentence as a linguistic object and a sentence as a carrier of conceptual meaning. In fact, our analysis shows that this dinstinction is instrumental for the intrinsic diagnosis of self-referential assertions. Buridan was apparently aware of this fact.

The world of concepts can be taken as a landscape whose forms and relationships stand available for recognition (conception). It is true, we ourselves participate more in this recognition than in the perception of sensory objects. But, nevertheless, we single out concepts as objects of our attention, observe distinctions and relations, as we would with objects of the sensory world. However, concepts do not appear to us in their essential structure without voluntary thought-activity. They do not drop into our consciousness by themselves. Hence we recognize a part of this landscape only as we are *actually* thinking.

It is important to notice here, that we are now dealing exclusively with concepts and not with assertions, sentences, etc. That is, the subjects and predicates of all propositions we are going to analyze are themselves conceptual.

Let us begin with an analysis of self-referential concepts, called *predicative* concepts. We follow Grelling/Nelson [1908] in their account of an idea going back to Russell (cf. also Finsler [1927b]).

> Every concept can either be applied to itself or not. The former shall be called "predicative", the latter "impredicative". (Examples of predicative concepts are: conceivable, abstract, consistent, invariant, as well as all other concepts which denote essential properties for the concepts themselves; In addition, many negative concepts are predicative, as for example, non-human, etc. The following are impredicative concepts: virtuous, green, and most of the everyday concepts.)[9]

Let us look at one of these predicative concepts more closely. The concept of abstractness means nothing else than the essential structure of all abstract objects. Let us call this structure "S_A". Hence S_A is the essential structure of all concepts which have the property of being abstract. Let us now single out a specific abstract concept as an object of our attention; the concept of the Riemann integral from calculus will do nicely. The Riemann integral, let us call it $S^{(i)}$, has an abstract structure, is an instance of S_A, therefore in the light of our previous considerations, we have

(12) $$S_A \rhd S^{(i)}.$$

On the left hand side is a conceptual category as before, namely the essential structure of abstractness, S_A. On the right hand side, however, there is no longer an object of the sensory world, as the

[9] Grelling/Nelson [1908: 60f.]. Translation by R. Ziegler/D. Booth

linguistic expression in (10'), but a conceptual entity. This conceptual entity, $S^{(i)}$, stands as a specific instance of the conceptual category S_A in the same manner as linguistic expressions were specific instances in our earlier analysis of self-referential assertions. In other words, (12) is a descriptive proposition about the specific concept $S^{(i)}$.

To obtain a self-referential concept of the type treated by Grelling/Nelson, we must now take S_A to be an instance of itself. Abstractness is itself abstract, or in symbols:

(13) $$S_A \rhd S_A \,.$$

Now, in this notation, the referrer and the referent seem to be one and the same. Thinking about this we realize that there is an actual distinction: The referrer and the referent arise differently. Were they undifferentiated, there would be no reference at all, in particular no self-reference. To interpret (13) consistently, the left side must be taken as a conceptual category and the right side as a specific instance of it.

Taking S_A as an instance of itself, is tantamount to assigning S_A a different quality beyond its conceptual structure or meaning. This structure appears as a part of the conceptual landscape, having its own ontological substance without which it has no existence. Any quality which we state *about* concepts (as, e. g., the abstractness of the Riemann integral) is not an essential part of their structural content, and hence not relevant for purely logical or mathematical considerations: But it is essential to their form of existence or appearance. The analysis of self-referential concepts shows that, in general, on cannot talk *about* concepts while denying them any kind of ontological quality: There would be nothing left to talk about.

Now let us turn to the concept of predicativity itself as described by Grelling/Nelson [1908]. In general, a concept C is called *predicative* or *self-referential,* if

(14) $$C \rhd C \,.$$

The distinction between the referrer and the referent is made explicit by our method of separating parameters:

(14') $$C^{(1)} \rhd C^{(2)}.$$

The concept of predicativity (or self-reference) contains an element not found in the concept of abstractness. The actual relationship between the conceptual category and one of its instances is the essential structural part of the concept of predicativity; it is not essential (in fact accidental) to the concept of abstractness. When Grelling/Nelson introduced the concept "predicative", they lead us to single out this relation. In fact, this relation is exactly what constitutes the concept of predicativity. Our intrinsic analysis therefore requires that we introduce a third parameter which expresses the fact that the relation between the constituents of the concept predicative, namely the referrer and the referent, is but a different representation of these constituents themselves:

$$C: C^{(1)} \triangleright C^{(2)}. \tag{15}$$

Without this third parameter C, we would only have a predicative (self-referential) concept and not the concept of predicativity itself.

Our analysis of predicative, or self-referential, concepts produces the same patterns here in (14) and (15) as were obtained in the analysis of self-referential (linguistic) assertions. It is truly remarkable that these patterns are alike; for in our analysis of the self-descriptive statement (10') we were moved by noticing a *perceptual* aspect in the statement. No actual perception arises out of the concept of predicativity. As we ponder predicative concepts we need to isolate aspects of them for our attention just as we need to distinguish the perceptual and conceptual components in self-referential assertions. From this observation we obtain two of our three parameters. These stand for *distinguishable* objects in the sense of section 2, where one has a property the other lacks. This distinction plays an important role in the following analysis of logical antinomies.

11. Diagnosis of logical antinomies

Having analyzed self-referential concepts in general, let us now turn to the analysis of a logical antinomy. The most direct of the logical antinomies is that which Grelling/Nelson [1908, 60f.] report as being from Russell. We shall continue the quotation begun in the previous section.

> The concept *impredicative* is itself either predicative or impredicative. Assume that it is predicative, it follows from its definition that it is impredicative. Assuming that it is impredicative, it does not apply to it itself, hence it would not be impredicative. Both assumptions lead to a contradiction.

The very concept of impredicativity is the focus of this antinomy. Any concept, C, does apply to itself, i. e. $C \triangleright C$, or does not appply to itself, $C \triangleright$ non-C. In the former case it is said to be predicative or self-referential; in the latter impredicative or non-self-referential. The question is, whether the concept impredicative, let us call it C_{im}, is predicative or not. The concept impredicative comprizes all concepts C which have the property of being impredicative, i. e. all concepts C with $C \triangleright \text{non-}C$.

The derivation of the antinomy using the present notation proceeds in the following way. In case C_{im} is actually predicative, we write $C_{im} \triangleright C_{im}$. But because of the meaning of C_{im}, we would then have: $C_{im} \triangleright \text{non-}C_{im}$. On the other hand, in case C_{im} actually is impredicative, one has $C_{im} \triangleright \text{non-}C_{im}$, so, taking in account the meaning of C_{im}, we have: $C_{im} \triangleright C_{im}$.

In deriving this contradiction, we begun with the assumption that $C_{im} \triangleright C_{im}$. This assumption is not merely formal, but has actual consequences. The very meaning of C_{im} is that of concepts being not self-referential. So the antinomical reasoning leads us to $C_{im} \triangleright \text{non-}C_{im}$. Were we to begin with the opposite assumption, $C_{im} \triangleright \text{non-}C_{im}$, a parallel chain of thought produces a contradiction too.

An intrinsic analysis of this argument leads us to separate parameters, so that we can go beyond the superficial pattern of the argument. The argument heads of with the assumption that C_{im} is actually impredicative, i. e. $C_{im} \triangleright C_{im}$. Right here, however, we have to distinguish between the conceptual content of C_{im} on the left and the concept of C_{im} as an object, a conceptual entity on the right, giving

$$C_{im}^{(1)} \triangleright C_{im}^{(2)}. \tag{16}$$

This expression says that $C_{im}^{(2)}$ is an *instance* of $C_{im}^{(1)}$, that is, the concept impredicative, taken as an *object* (conceptual entity), has to satisfy its own conceptual content. Hence we have

$$C_{im}^{(2)} \triangleright \text{non-}C_{im}^{(2)}. \tag{17}$$

The things denoted by $C_{im}^{(2)}$ in (16) and (17) are distinguishable objects in the sense of section 2. In (16), $C_{im}^{(2)}$ is a specific conceptual object, which is an instance of the conceptual content $C_{im}^{(1)}$, whereas in (17), $C_{im}^{(2)}$ itself is a conceptual content of which $\text{non-}C_{im}^{(2)}$ is an instance.

Conjoining (16) and (17) gives a contradiction if and only if $C_{im}^{(1)}$ and $C_{im}^{(2)}$ are identified. In identifying these parameters, however, we loose the very distinction which allowed us to make the transition from (16) to (17) in our chain of reasoning.

A similar argument applies to the chain of reasoning starting with the opposite assumption, $C_{im} \triangleright \text{non-}C_{im}$. We conclude that this antinomy is equivalent to neglecting the distinction between the content of a concept and the concept as a substantial object (conceptual entity or instance) in itself.

One might think that this distinction amounts to a typed response. However, type theory does not explain or actually solve an antinomy, it simply avoids them. In contrast, our approach goes back to the roots of logical antinomies and thus reveals the reason why they arise in the first place.

This paradox concerning the concept impredicative seems to be an antinomy only to the thinker who neglects the principle of identity by confusing the content of a concept with the concept as a substantial object (not from the sense-perceptual reality, of course), i.e. the concept as an *instance* of another (here, the same) concept. In other words, a pragmatic contradiction (in our earlier terminology, distinction) is projected onto the field of conceptual relations, producing what appears there to be an analytical contradiction.

In his discussion of the antinomy concerning the concept "impredicative", Finsler [1927b] stresses the fact that we cannot expect circular or self-referential definitions to be satisfiable in any case. There are exceptions, and the definition of "impredicative" has the property that it cannot be applied to itself without producing a contradiction. Therefore, applied to itself, the concept "impredicative" has no meaning.

Russell's set theoretic paradox arises from the antinomy of the concept "impredicative" if we switch to the extensional point of view. Sets which do contain themselves as elements ("predicative sets") are not well-founded. Sets which do not contain themselves ("impredicative sets") are well-founded; they are also called "normal". The set of *all* normal sets is Russell's set R. If R is normal, it does not contain itself as an element hence R is not normal. Conversely, if R is not normal, it does contain itself, hence it is normal. This is the antinomy.

Finsler [1926b], [1927b] draws from this situation the same conclusion as above: Self-referential definitions need not be satisfiable.

Conclusion

All references involve the distinction between a refering part (referrer) and an object which is referred to (referent): This distinction is easily forgotten where self-reference is concerned. But in reality it is still there. Were we to neglect this distinction, there would be no actual reference at all, merely an identity. In deriving a contradiction, using one of the self-referential patterns discussed earlier in this paper one actually recognizes first of all the distinction between the referrer and the referent. The given distinct objects have to be recognized as real. But, in the next stage of the argument, the distinction is dropped, an identity is assumed, and the illusion of a contradiction is produced. Hence, the antinomical argument produces an illusory contradiction from a real distinction, based upon a violation of the principle of identity.

Within the antinomical patterns, there lurks a contradiction waiting to catch us if we neglect the deeper aspects of self-reference. The need for intrinsic analysis of the deeper aspects of reference exist in any instance of self-reference, quite apart from whether an antinomy arises or not. The antinomies merely make us aware of something that is present in all self-reference. Hence the heart of the problem is the structure of self-reference.

When there is self-reference, the referrer is a concept and the referent is an observed instance, so observation is necessary for self-reference. For self-referential assertions, like the truth-teller or the Liar, which involve actual linguistic expressions, the observational part of the reference is directly related to sensory perception. For self-referential concepts, however, the principle is the same, but the observation is something deeper and more elusive, an observation *within* the conceptual realm. This is a necessary consequence of our accepting the existence of self-referential concepts. This observation, a primary experience by its very nature, reveals a substantial component, namely some kind of ontological basis of conceptual entities.

Thus it turns out that the question of whether concepts have an existence independent from our mind is not a purely theoretical one, subject only to individual beliefs. This would mean also that this question cannot be decided by deductive reasoning from some "well-known" principles. On the contrary, as we have shown, what is needed is the observation and analysis of our own thinking process. The basic observational facts and experiences must be drawn from carefully designed mental experiments. In this paper, one exceptionally rich example was presented and analyzed, namely the concept of self-reference and some of its manifestations.

Someone who attempts to deny this observational (but *not* sense-perceptual) component of concepts as an objection to these views, must exclude all self-referential concepts from his considerations. In fact, this was done by Russell [1908], who explicitly eliminated all self-referential structures. However, this cannot be done consistently, as was shown conclusively by Várdy [1979]. Suppose, self-reference to be forbidden: No assertion or concept applies to itself. Obviously, this assertion has a self-referential structure and denies in its propositional content all self-reference, in particular its own. This leads to an antinomy.

An intrinsic analysis of semantic and logical antinomies shows that they, in fact, rest upon the same structure. In deriving the contradiction, one has to violate the principle of identity, while identifying the referrer and the referent. Therefore, from this point of view, semantic and logical antinomies are similar. However, they differ in the realm from which the referent is drawn. The referents in the semantic antinomies are drawn from perception. The referents of the logical antinomies, on the other hand, are drawn from the realm of pure concepts, hence from conception. It is obvious, in the case of semantic antinomies, that the specific perceptual reference transcends logical forms. Even the logical antinomies contain specific references which are not part of formal logic, though they remain accessible to human reason. Pure logic, after all, concerns itself with *possible* objects: The distinctions that underlie the antinomies ultimately refer to *actual* objects. In other words, our intrinsic analysis of logical antinomies concerns itself with the transition from conceptual propositions to descriptive propositions within the realm of pure concepts. Self-reference is the key which opens our minds to the experience of the ontological status of concepts.

Acknowledgements

This paper is the result of many discussion I had with Werner A. Moser and owes much to his clear, crisp and yet very experiential approach to systematic philosophy. For a more elaborate discussion of some of the topics developed here, see the book Ziegler [1995]. Additional helpful comments came from Bernd Gerold, Thomas Kesselring, and Thomas Meyer. Discussions with David Booth while translating this paper together resulted in many improvements, both conceptually and linguistically.

Finsler Set Theory: Platonism and Circularity

Translation of Paul Finsler's papers on set theory with introductory comments

Edited by David Booth
and Renatus Ziegler

Birkhäuser Verlag
Basel · Boston · Berlin

Editors' Addresses

David Booth
Three Fold Foundation 307
Hungry Hollow Road
Chestnut Ridge
New York 10977
USA

Renatus Ziegler
Mathematisch-Astronomische Sektion am Goetheanum
4143 Dornach
Switzerland

Mathematics Subject Classification (1991): 03A05, 03E05, 03E20, 03E30, 04A20

A CIP catalogue record for this book is available from the Library of Congress, Washington D.C., USA

Deutsche Bibliothek Cataloging-in-Publication Data

Finsler set theory: Platonism and circularity : translation of Paul Finsler's papers on set theory with introductory comments / ed. by David Booth and Renatus Ziegler. - Basel ; Boston ; Berlin : Birkhäuser, 1996
ISBN 3-7643-5400-3 (Basel...)
ISBN o-8176-5400-3 (Boston)
NE: B ooth, David [Hrsg.]; Finsler, Paul

Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Cover design: Markus Etterich, Basel
Cover illustration: Photo of Paul Finsler by courtesy of Regula Lips-Finsler, handwritten manuscript of Paul Finsler by courtesy of the Mathematical Institute, University at Zürich
Printed in Germany
ISBN 3-7643-5400-3
ISBN 0-8176-5400-3

9 8 7 6 5 4 3 2 1

Contents