ON SOME NON-PARADOXICAL SELFREFERENTIAL SENTENCES

Matthias VARGA VON KIBÉD

Consider the following sentences:

(A) Sentence (A) is false. ["The liar"]
(B) Sentence (B) is true. ["The truth-teller"]
(C) Sentence (C) is neither true nor false.
(D) Sentence (D) is either true or false.

Sentence (A) is a familiar version of the venerable liar paradox; sentence (B) is the socalled truth-teller. In [K] A. Kukla considers a peculiarity of the truth-teller sentence closely related to the problems of the liar paradox by using an argument of Mortensen and Priest in [M&P]. By extending this kind of discussion to sentences like (C) and (D), Kukla gets the possibly somewhat surprising (informal) result, that in contrast to the paradoxicality of (A) and (the paradoxical flavour of certain arguments about) (B), sentence (C) is false and sentence (D) is true, both on a priori grounds. Kukla stresses, that a proposed solution to the liar paradox should similarily elucidate the status of (B), (C) and (D) in order to be acceptable. I agree with this postulate of Kukla, but would like to give a hint for a formal frame, in which the arguments of Kukla (and the for his argument relevant parts of Mortensen and Priest [M&P]) can be represented adequately. Moreover, some weak points of these expositions will thereby become visible. Sentence (C) will be seen to retain some paradoxical features overlooked by Kukla.

For my approach I will use as formal frame M an amalgamation of Smullyans selfreferential ([1]) language SELF and Blaus logic of reflection ([2]) LR. Smullyans SELF gives a minimal frame for representing selfreference ([3]), while a weak subsystem of Blau's LR seems to be at pre-

([1]) First expounded by Smullyan in [S]; a short exposition is to be found in Manin [M], detailed comments in Stegmüller and Varga [S&V].

([2]) Cf. Blau [B]; the version given here is to be found in more detail in Varga [V].

([3]) As argued in [S&V].

sent one of the most convincing proposals for a (non-eliminative) solution of the liar paradox including an analysis of selfreferential sentences like (B)-(D) above. ([4])

In *M* we have as symbols the one place predicates "W", "F", "D", "I" of truth, falsity, definiteness (i.e. truth-or-falsity) and indefiniteness (i.e. neither-truth-nor-falsity). The extension of these predicates will have to be suitable sets of expressions of *M*. Besides, *M* has the connective "–" for negation, quotes (of the object language) "⌜⌝" and the "norm name forming" operator "E". In order to understand the use of "E" we consider expressions of the form "⌜e⌝" as *M*-names ([5]) of e, call "e ⌜e⌝" *the norm of* e and let for any *M*-name e′ of e "Ee′" be a *M*-name ([6]) of the norm of e. Thus "E" functions as a sort of diagonal operator making it possible to represent syntactically selfreferential statements as we will see. *M*-expressions are arbitrary sequences of the symbols introduced above. ([7])

Formulas of *M* have the form

(–) Pn(e)

that is a *M*-predicate, possibly preceded by "–" followed by some *M*-name n(e) of an *M*-expression e.

We now give some very simple semantical rules for evaluation of *M*-

([4]) Why I think that one should favour Blau's LR over the better known ideas of Kripke in his "Outline of a Theory of Truth" [Kr], Martin and Woodruff's [M&W], as well as the less known but related work of Kind [Ki] is beyond the reach of this short essay; suffice it to say that Blau's LR seems to give a more convincing motivation for his semantical rules (better in accord with the supervenience of semantics, cf. Kremer [Kre]), a much more comprehensive framework for the integration of metalinguistic features in the object language and less technical problems with ordinal arithmetic than other systems like those cited above or those of Gupta, Herzberger, van Fraassen, Yablo, Skyrmes and others to be found in R.L. Martin's [M1] and [M2] – at least with respect to the analysis of the semantical paradoxes (see also [V]). Kukla does not refer to any of these systems; therefore I will confine myself here to the exposition of *M* as a minimal frame for analyzing sentences (A) - (D) and Kukla's arguments about them.

([5]) More exactly: of the form "⌜" ^e^ "⌝", with "^" as the sign of concatenation.

([6]) More exactly "Ee′" and similar expressions should be given with Quine's quasi-quotation.

([7]) Care is to be taken in some way that quotes can be read unambiguously; some of the many ways to achieve this are explained in [V].

formulas. ([8]) The central idea is to use ordinal indices for the interpretation function ([9]). This way we are able to represent the typical forms of arguments about selfreferential statements. The semantics of *M* are most easily understood by using it on some selfreferential example. Let therefore sentence (A) be formally represented in *M* as

(A′) FE ⌜FE ⌝ .

To see that according to the given semantics (A′) represents (A), remember that F is to be interpreted as predicate for falsity of *M*-expressions and "E ⌜FE ⌝" is the name of the norm of the *M*-expression named by "⌜FE ⌝", i.e. (A′) itself. Thus (A′) "says" of itself that it is false.

Now how is (A′) to be treated semantically? The usual argument about (A) first states, that if (A) is true, then what it says has to be the case, and therefore, as it says that it is false it would have to be false, thus it can't be true. Similarly if (A) were false analogously it could be argued for its truth. Thus (A) is neither true nor false. ("Level O") But we now can use this result as a basis for saying that this contradicts what A says, and thus (A) is false. ("Level 1") But this is what (A) says and therefore (A) is true. ("Level 2") But .... This "oscillating" property of the liar has often been expounded. We now give it a preformal version in the semantics of *M*.

Write $|S|^n$ for the truth value of a *M*-sentence S on level n, where the level of an argument is similar to those hinted at just before for the liar. Then ([10])

$$|(A')|^0 = W \text{ implies } |(A')|^0 = F$$

and vice versa. Thus if we want to avoid a contradiction we should conclude, that (A′) on level 0 has no classical truth value. But as (A) says that it is false, i.e. has a classical truth value, we should evaluate (A′) on the next level 1 as false:

([8]) As we treat weak fragments of systems of sufficient descriptive richness, we can, by hinting at their existence as natural enlargements of *M* argue, that our semantics for *M* does not just work because of the minimality of the chosen frame of exposition.

([9]) For the aims of this essay, transfinite ordinals are nearly irrelevant. Thus the indices may be identified with natural numbers.

([10]) "W", "F" are taken for the classical truth values "true" and "false"; "O" will be used for the non-classical case. For a fully formalized version vide [V] and [B].

$$|(A')|^1 = F .$$

Now this is what (A) says, thus on level 2 we should have

$$|(A')|^2 = W .$$

By natural generalization of this kind of argument we find the oscillation of truth values, based in the non-classical case of level 0, closely paralleling the informal argument. A formal semantics of course has to include these inferences (about the values of (A') on different levels) in its mechanics. This is technically exploited e.g. in [B] and [V]. The circular argument for level 0 is to be seen then formally as to exhaust all possibilities for giving a classical truth value to (A') on level 0. This fact is then sufficient reason for evaluating (A') as incorrect falsity statement about the level 0 value of (A'). To get an idea about the type of semantic rules to be used, this evaluation of (A') on the basis of

$$|(A')|^0 = O \text{ ("O" representing the non-classical case)}$$

proceeds by the following rule

(R) $|F\ulcorner S\urcorner|^{n+1} = W$ iff $|S|^n = F$, and $= F$ otherwise ([11]).

Thus we have characterized the liar by the sequence

O, F, W, F, W, F, ...

Similarily we get for the truth-teller (B), formalized in *M* as

(B') $WE\ulcorner WE\urcorner$

the sequence

O, F, F, F, ... ;

thus the symmetrical argument given by many authors and repeated in Kukla [K] "there do not seem to be any arguments that establish its (i.e. the truth-tellers) truth or falsehood aprioristically" looses some of its naturalness. ([12])

Mortensen and Priest in [M&P] had already given a simple argument,

([11]) And analoguous rules for the predicates "W", "I", "D"; e.g.

(R') $|W\ulcorner S\urcorner|^{n+1} = W$ iff $|S|^n = W$; and $= F$ otherwise.

([12]) Cf. with more details [V2].

that the truth-teller (B) must be either true or false. For lets assume the contrary, i.e. let (B) be neither true nor false. Then this contradicts what (B) says of itself, i.e. that (B) is true. Therefore on this assumption we infer (B)'s falsity, but this contradicts the assumption. Thus we have arrived by reductio ad absurdum at the result, that the truth-teller has to have a classical truth value. By our considerations above the question, termed inpenetrable by Kukla and left unanswered also by Mortensen and Priest, which is the classical truth value of the truth-teller, finds a definite answer. ([13]) The non-classical value O for the truth-teller on level 0 still reflects the old symmetrical intuition, while the final stability of the truth value sequence in the classical value F represents our solution to the problem posed by Mortensen and Priest.

Now we can give a similar analysis for Kukla's sentences (C) and (D). For (C), formalized as

(C′) IE ⌜IE ⌝

we get the sequence

O, W, F, F, F, ...

To see this, just note that on level 0 we won't find a classical truth value, as the intended semantical rules will trace back the truth value of a sentence Pa to the question, whether the extension | a | has the appropriate property for the interpretation of P; but in the case of (C′) the extension of the argument "E ⌜IE ⌝" is just (C′) itself, thus leading to a regressus ad infinitum. On level 1 we therefore have to evaluate the I-predication for (C′) as true, because (C′) indeed has no classical truth value on level 0. At higher levels then (C′) will be continuously evaluated as false, because of the classical truth value on level 1.

Now let's have a look at Kukla's argument. He supposes that (C) is neither true nor false and infers, as this is what (C) says that then (C) would have to be true, contradicting the assumption. Thus (C) is not neither true nor false, i.e. either true or false. Which one? If (C) were

([13]) Which was difficult to see for many authors because they overemphasized the symmetrical situation for the truth-teller's evaluation which arises with hypothetical reasoning: if (B) were true, well then it's just true; and similarly if it were false no problem arises. This hypothetical reasoning in our approach is embedded in the non-classical level 0 value for (B′).

true, from what (C) claims would follow its non-truth. Therefore, so Kukla argues, (C) is false.

What should we say about this line of reasoning on the background of the frame of reconstruction given above?

(i) Kukla's argument leads to a similar result as our reconstruction, but
(ii) Kukla implicitly assumes, that there are no non-classical cases to be considered, because otherwise he could not infer the falsity of (C) from its non-truth. By this
(iii) the level 1 perspective, where (C)'s truth "before" its finally stabilized classical truth value "false" is overlooked by Kukla.

Kukla's argument for the truth of (D) runs exactly dual to the above repeated argument for the falsity of (C). Consider a formal version for (D):

(D′) DE ⌜DE ⌝.

This will be evaluated by the truth value sequence

O, F, W, W, W, ...

which is exactly dual to the sequence for (C′) ([14]).

But in contrast to Kukla we see that two aspects of paradoxicality remain, which are invisible because of the overlooked level 1 perspective of (III):

(a) We have no classical truth value at level 0
(b) There are "points of view", formally represented as "levels of evaluation" with conflicting classical truth values for (C) as well as for (D).

The non-paradoxicality of (C) and (D), which Kukla along the lines of Mortensen and Priest wanted to stress, is to be found in the "convergence" of the truth value sequences with classical "limits" in both cases. Finally the informal argument of Kukla about (C) and (D) in some sense can't be accepted in his setting, because, as he says himself "the same kind of reasoning leads to a contradiction" when applied to the liar (A). We thus have given here a formal background where the intuitive validity

([14]) Let X^0 denote the dual truth value for X and set $O^0 := O$, $W^0 := F$ and $F^0 := W$.

of much of such an informal argument for (C) and (D) can be formally motivated.

Let's return for a moment to Mortensen and Priest's argument for the classicality of the truth-tellers truth value. The statement,

(E) The truth-teller (B) has a classical truth value

(E′) D ⌜WE ⌜WE ⌝ ⌝;

we leave it as an exercise to the reader to verify by an argument along the lines given above that we get the sequence

O, F, W, W, W, ...

of truth values for (E′). Inspecting the informal argument given above shows that here the level 1 perspective was not overlooked by Mortensen and Priest. And, by the way, we get the maybe somewhat surprising equivalence of (D) and (E) for all levels of evaluation of their formal versions!

Universität München M. VARGA VON KIBÉD
Seminar für Philosophie,
Logik und Wissenschaftstheorie

BIBLIOGRAPHY

[B] U. Blau, "Die Logik der Unbestimmtheiten und Paradoxien", *Erkenntnis* 22 (1985), 369-459.

[Ki] W. Kindt, "Über Sprachen mit Wahrheitsprädikat", *Sprachdynamik und Sprachstruktur*, eds. Habel, Kanngiesser, Tübingen (1976).

[Kre] M. Kremer, "Kripke and the Logic of Truth", *Journal of Philosophical Logic* 17 (1988), 225-278.

[Kr] S.A. Kripke, "Outline of a Theory of Truth", *Journal of Philosophy* 72 (1975), 690-716.

[K] A. Kukla, "Two Surprisingly Non-Paradoxical Sentences", *Logique et Analyse* 28 (1985), 109-110.

[M1] R.L. Martin (ed.), *The Paradox of the Liar*, New Haven and London (1970).

[M2] R.L. Martin (ed.), *Recent Essays on Truth and the Liar Paradox*, Oxford, New York (1984).

[M] Yu.I. Manin, *A Course in Mathematical Logic*, Berlin, Heidelberg, New York (1976).

[M&P] C. Mortensen and G. Priest, "The Truth Teller Paradox", *Logique et Analyse* 24 (1981), 281-288.

[M&W] R.L. Martin and P.W. Woodruff, "On Representing 'True-in-L' in L", *Philosophia* 5 (1975), 213-217.

[S] R.M. Smullyan, "Languages in Which Self-Reference is Possible", *Journal of Symbolic Logic* 22 (1957), 55-67.

[S&V] W. Stegmüller und M. Varga von Kibéd, *Strukturtypen der Logik*, Berlin, Heidelberg, New York (1984).

[V] M. Varga von Kibéd, *Wahrheit, Selbstreferenz und Reflexion*, Ph.D. Diss., München (1987).

[V2] M. Varga von Kibéd, "Symmetrische und Asymmetrische Auffassungen vom Truth-Teller", forthcoming in *Analytica*.