The Elimination of Self-Reference: Generalized Yablo-Series and the Theory of Truth

Although it was traditionally thought that self-reference is a crucial ingredient of semantic paradoxes (e.g. *This sentence is false)*, Yablo (e.g. 2004) showed that this is not so by displaying an infinite series of non-self-referential sentences which, taken together, are paradoxical. Let us write $\{<s(\mathbf{k}), F_k>: k\geq 0\}$ for a denumerable set of pairs whose second coordinate is a sentence named by the first coordinate (we call such sets 'naming relations'). It can be shown that each of the following naming relations is paradoxical (*Tr* is interpreted as the truth predicate; **i**, k, k' are integer-denoting):

(1) a. Universal Liar: $S_\forall := \{<s(\mathbf{i}), \forall k\ (k>\mathbf{i} \Rightarrow \neg Tr(s(k)))>: i\geq 0\}$[1]

b. Existential Liar: $S_\exists := \{<s(\mathbf{i}), \exists k\ (k>\mathbf{i} \Rightarrow \neg Tr(s(k)))>: i\geq 0\}$

c. Almost Universal Liar: $S_{AA} := \{<s(\mathbf{i}), \exists k\ (k> \mathbf{i} \wedge \forall k'\ (k'>k \Rightarrow \neg Tr(s(k'))))> : i\geq 0\}$

We generalize Yablo's construction along two dimensions: (i) First, we investigate the behavior of Yablo-style series of the form $\{<s(\mathbf{i}), [Qk: k>\mathbf{i}]\ Tr(s(k))>: i\geq 0\}$, for some generalized quantifier Q. We show that for any Q that satisfies certain natural properties, all the sentences in the series must have the same value. We derive a characterization of those values of Q for which the series is paradoxical. (ii) Second, we show Yablo's results are a special case of a much more general phenomenon: given certain assumptions, *any semantic phenomenon that involves self-reference can be 'imitated' without self-reference*. The result is proven for Kripke's Theory of Truth with the Strong Kleene Logic (Kripke 1975).

1. Yablo-Series with Generalized Quantifiers

We naming relations of the form in (2), where Q is a binary generalized quantifier (e.g. *some, most, no, all, an odd number of,* etc.) which satisfies the properties of Permutation Invariance, Extension and Conservativity. For special values of Q we obtain versions of Yablo's paradox:

(2) $S_Q=\{<s(\mathbf{i}), [Qk: k>\mathbf{i}]\ s(k)>: i\geq 0\}$

a. For Q=*No*, S_Q is the Universal Liar.

b. For Q=*Not all*, S_Q is the Existential Liar.

c. For Q=*All but a finite number of,* S_Q is the Almost Universal Liar.

(3) A relations R of subsets of E satisfies:

a. Permutation Invariance just in case for all E, for any permutation π of E, for all X, Y $\subseteq$E, $R_E(X, Y)$ iff $R_E(\pi(X), \pi(Y))$

b. Extension iff: for any X, Y, E, E', if X, Y$\subseteq$ E and X, Y$\subseteq$ E', then $R_E(X, Y)$ iff $R_{E'}(X, Y)$

c. Conservativity iff for all X, Y, E: $R_E(X, Y)$ iff $R_E(X, X\cap Y)$

In a bivalent logic, a generalized quantifier Q that satisfies the conditions in (3) is defined by its 'tree of numbers' $\underline{Q}^2$, which is a a function from pairs of numbers (including ∞) to truth values such that: for any formulas F, F' with extensions $\underline{F}$ and $\underline{F}'$, *Qx F F'* is true (in a bivalent system) iff $\underline{Q}^2(<|\underline{F}-\underline{F}'|, |\underline{F}\cap\underline{F}'|>)=1$(van Benthem 1986). We study S_Q in any n-value logic which is 'reasonable', in the sense that the semantics of the quantifiers satisfies a generalization of the tree of numbers:

(4) An n-valued logic with truth values in E is *reasonable* just in case:

If for any assignment function F has a classical value, then for any generalized quantifier Q, the value of a closed formula *[Qk: F]F'* only depends on

$(|\{d\in D: [[F]]^{k\to d}=1\}\cap\{d\in D: [[F']]^{k\to d}=e\}|)_{e\in E}$.

We show that if a reasonable compositional logic has a finite number of truth values, *all the sentences in S_Q must have the same truth value*. We derive a characterization of those values of Q for which S_Q is paradoxical in a bivalent or trivalent system:

(5) Let Q be a binary generalized quantifiers satisfying Permutation Invariance, Extension and Conservativity. Then:

a. A binary valuation can be found in which S_Q has the value *true* iff $\underline{Q}^2(<0, \infty>)=1$

b. A binary valuation can be found in which S_Q has the value *false* iff $\underline{Q}^2(<\infty,0>)=0$

c. S_Q is paradoxical iff no binary valuation can be found in which S_Q has the value *true* and no binary valuation can be found in which S_Q has the value *false,* iff $\underline{Q}^2(<0, \infty>)=0$ and $\underline{Q}^2(<\infty,0>)=1$

---

[1] To see that this series is paradoxical: (i) Suppose all sentences are false. Then what each of them says is true - contradiction. (ii) Suppose that *s(**i**)* is true. Then *s(**i+1**), s(**i+2**), s(**i+3**),* etc. are false - which should make *s(**i+1**)* true!

2. Elimination of Self-Reference

Cook 2004 considers a primitive setting in which infinite conjunction replaces quantification over sentences, and shows that in his system every paradox that involves self-reference can be 'unwinded' to give rise to a Yablo-style paradox without self-reference. We generalize Yablo's and Cook's constructions by showing that under certain conditions, *a language with self-reference can be translated into a self-reference-free fragment of a language with quantification over sentences.* The analysis is framed within Kripke's theory of truth, so as to apply not just to purely logical paradoxes, as in Cook's framework, but also to 'empirical' paradoxes (e.g. *Every statement made by Nixon about Watergate is false;* as uttered by Nixon, this statement may or may not be paradoxical depending on some empirical facts).

We start from a classical language L without quantifiers, to which we add a truth predicate *Tr* whose interpretation is partial (trivalent); we call the resulting language L', and specify a bijective naming relation N over L' (i.e. each sentence of L' has exactly one name). For each pair $<\underline{s}, s>$ of N (where $\underline{s}$ is a term denoting the formula s), we define a series of translations $\{<\underline{s}(\mathbf{k}), h_k(s)>: k\geq 0\}$ in a quantificational language L* that extends L (we also write: $h_k(<\underline{s}, s>)=<\underline{s}(\mathbf{k}), h_k(s)>$). We fix a classical interpretation I for L, and restrict attention to interpretations of L' and L* that extend I and are fixed points in the sense of Kripke 1975. It can be shown that:

P1. None of the translations is self-referential, i.e. for no k is $h_k(s)$ self-referential.

P2. In any fixed point I* of L* compatible[2] with N, all the translations of a given formula s of L have the same value according to I*, i.e. for all k, $k'\geq 0$, $I^*(h_k(s))=I^*(h_{k'}(s))$.

P3. (a) for every fixed point I' of L' compatible with N there is a fixed point I* of L* compatible with h[N] such that for each sentence s of L', $I'(s)=I^*(h_k(s))$ [notation: $h[N] := \{h_k(<\underline{s}, s>: <\underline{s}, s>\in N \wedge k\geq 0\}$]. Conversely, (b) for every fixed point I* of L* compatible with h[N] there is a fixed point I' of L' compatible with N such that for each sentence s of L', $I'(s)=I^*(h_k(s))$.

The translation procedure h is defined in (6) and illustrated in (7)-(10):

(6) Let $[Qk': k'>\mathbf{k}]F$ abbreviate: $\exists k''\ (k''>\mathbf{k} \wedge \forall k'\ (k'\geq k'' \rightarrow F))$.
If $<\underline{s}, s>\in N$, $h_k<\underline{s}, s> = <\underline{s}(\mathbf{k}), [Qk': k'>k][s]_{k'}>$
where $[s]_{k'}$ is the result of substituting each occurrence of the form *Tr(c)* in s with *Tr(c(k'))*.

(7) Suppose that $<c_1, P^0_1>\in N$, where P^0_1 is an atomic proposition. Then:
$h_k<c_1, P^0_1> = <c_1(\mathbf{k}), [Qk': k'>\mathbf{k}]\ P^0_1>$
Note that the quantification is vacuous, since P^0_1 does not contain any variables. For any interpretation I for L and for any interpretations I' and I* which extend I to L' and L* respectively, for each $k\geq 0$,
$I^*(h_k(P^0_1))=I^*([Qk': k'>k]\ P^0_1)= I^*(P^0_1)=I(P^0_1)=I'(P^0_1)$

(8) Suppose that $<c_2, Tr(c_1)> \in N$, with c_1 as in (7).
$h_k<c_2, Tr(c_1)> =<c_2(\mathbf{k}), [Qk': k'>\mathbf{k}]Tr(c_1(k'))>$

(9) Suppose that $<c_3, \neg Tr(c_3)>\in N$.
$h_k<c_3, \neg Tr(c_3)>= <c_3(\mathbf{k}), [Qk': k'>\mathbf{k}]\ \neg Tr(c_3(k'))>$
It is clear that $\{<c_3, \neg Tr(c_3)>\}$ and $\{<c_3(\mathbf{k}), [Qk': k'>\mathbf{k}]\ \neg Tr(c_3(k'))>: k\geq 0\}$ are both Liar-like: the former is the simple Liar, and the latter is the Almost Universal Liar.

(10) Suppose that $<c_4, Tr(c_4)> \in N$.
$h_k<c_4, Tr(c_4)>= <c_4(\mathbf{k}), [Qk': k'>\mathbf{k}]Tr(c_4(k'))>$
$\{<c_4, Tr(c_4)>\}$ is the 'Truth-Teller', and $\{<c_4(\mathbf{k}), [Qk': k'>\mathbf{k}]Tr(c_4(k'))>: k\geq 0\}$ is an infinite Truth-Teller: all sentences in the series must have the same truth value, but it may be chosen arbitrarily.

We consider alternative values of Q and characterize those that can be used in the translation:

(11) Q can be used in the translation h if and only if for all finite $i\geq 0$, $\underline{Q}^2(<\infty, i>)=0$ and $\underline{Q}^2(<i, \infty>)=1$
In particular, we show that when the latter condition fails, Property **P2** fails to hold.

When we restrict attention to infinite universes, this gives only two quantifiers: Q=*all but finitely many* (which is, in effect, the quantifier used in (6)) and Q=*infinitely many*.

References: Cook, R. 2004. 'Patterns of Paradox', *Journal of Symbolic Logic,* 69, 3, 767-774; Kripke, S. 1975. 'Outline of a Theory of Truth', *Journal of Philosophy* 72: 690-716; van Benthem, J. 1986. *Essays in Logical Semantics*, Reidel, Dordrecht; Yablo, S. 2004. 'Circularity and Paradox', in *Self-Reference,* CSLI.

[2] An interpretation I is compatible with a naming relation N if for each $<s, F>\in N$, $I(s)=F$.