Logical paradoxes

David Pierce

September 23, 2013

Mathematics Department
Mimar Sinan Fine Arts University
dpierce@msgsu.edu.tr
http://mat.msgsu.edu.tr/~dpierce/

These notes were prepared in the fall of 2013 to distil the common essence of certain paradoxes, from the Russell Paradox to Gödel's Incompleteness Theorem. Using the work of R. G. Collingwood (1889–1943), the last section observes that a science purporting to study the whole world must also study itself *as* a science. LaTeX file last compiled, December 1, 2014.

Contents

1 Introduction

The barber paradox is that there can be no male barber who shaves every man that does not shave himself. If there were such a barber, he would shave himself if and only if he did not.

The Russell paradox, in its original form [6], is that *to be a predicate that cannot be predicated of itself* is not a predicate. If it were a predicate, then it could be predicated of itself if and only if it could not. Similarly there is no class consisting of the classes that do not belong to themselves.

These paradoxes are instances of the following.

Theorem 1. *Let R be a binary relation on some domain of individuals. Then there is no individual a in that domain such that, for all individuals x in the domain,*

$$a\,R\,x \iff \neg\, x\,R\,x.$$

Proof. Replacing x with a yields a contradiction. $\square$

2 Tarski's Theorem

Tarski's Undefinability Theorem [7, Thm. I, §5, p. 247] can now be understood as follows. We take a *logic* to be a formal way of studying some domain of individuals. The logic has *formulas,* each of which may have free variables. The variables range over the domain being studied. A formula with no free variables is a *sentence.* A sentence is true or false. A formula with one free variable is a *singulary* formula. If φ is singulary, and a is an individual, then $\varphi(a)$ is a sentence: it is the result of replacing each free occurrence of the free variable of φ with a name for a.

We suppose each formula φ has a *code,* denoted by $\ulcorner\varphi\urcorner$. This code is an individual in the domain of our logic. In Theorem 1, take R to be relation that obtains between singulary formulas φ and ψ just in case the sentence $\varphi(\ulcorner\psi\urcorner)$ is true. That is,

$$\varphi \mathrel{R} \psi \iff \varphi(\ulcorner\psi\urcorner).$$

Then there is no singulary formula θ such that, for all singulary formulas φ,

$$\theta(\ulcorner\varphi\urcorner) \iff \neg\varphi(\ulcorner\varphi\urcorner).$$

Now suppose that, for every singulary formula ψ, there is a singulary formula ψ^* such that, for all singulary formulas φ,

$$\psi^*(\ulcorner\varphi\urcorner) \iff \psi(\ulcorner\varphi(\ulcorner\varphi\urcorner)\urcorner). \tag{1}$$

Then there is no singulary formula ρ such that, for all sentences σ,

$$\rho(\ulcorner\sigma\urcorner) \iff \sigma. \tag{2}$$

That is, there is no formula defining the collection of codes of true sentences. For, if there were such a formula ρ, then

$$\neg\rho^*(\ulcorner\varphi\urcorner) \iff \neg\rho(\ulcorner\varphi(\ulcorner\varphi\urcorner)\urcorner) \iff \neg\varphi(\ulcorner\varphi\urcorner),$$

that is, $\neg\rho^*$ would be a formula θ as above. Here we understand $\neg\rho^*$ to be $\neg(\rho^*)$, although this is equivalent to $(\neg\rho)^*$.

Another way to understand this is to note a special case of (1):

$$\psi^*(\ulcorner\psi^*\urcorner) \iff \psi(\ulcorner\psi^*(\ulcorner\psi^*\urcorner)\urcorner).$$

Thus, for every singulary ψ, there is a sentence $\tilde{\psi}$ such that

$$\tilde{\psi} \iff \psi(\ulcorner\tilde{\psi}\urcorner).$$

This is the "Diagonal Lemma" (see the article of that name in Wikipedia). In particular, if (2) holds, then

$$\rho(\ulcorner\widetilde{\neg\rho}\urcorner) \iff \widetilde{\neg\rho} \iff \neg\rho(\ulcorner\widetilde{\neg\rho}\urcorner),$$

which is a contradiction.

3 Gödel's Theorem

Gödel's Incompleteness Theorem [5], in its most basic form, relies on a variant of Theorem 1.

Theorem 2. *Suppose Q and R are binary relations on some domain, and R includes Q, that is,*

$$x\,Q\,y \implies x\,R\,y.$$

Suppose there is an individual a such that

$$a\,R\,x \iff \neg\, x\,Q\,x.$$

Then Q and R are not the same relation, and indeed

$$\neg\, a\,Q\,a \;\;\&\;\; a\,R\,a.$$

Proof. Since

$$a\,Q\,a \implies a\,R\,a \implies \neg\, a\,Q\,a,$$

it follows that $\neg\, a\,Q\,a$ and therefore $a\,R\,a$. □

In the logic that we considered above, we now assume that there is a notion of *provability* of sentences. All provable sentences are true. In addition to defining R as before, we now let Q be the relation on singulary formulas given by

$$\varphi\,Q\,\psi \iff \varphi(\ulcorner\psi\urcorner) \text{ is provable.}$$

We assume there is a formula χ such that, for all sentences σ,

$$\chi(\ulcorner\sigma\urcorner) \iff \sigma \text{ is provable.}$$

Then we have

$$\chi^*(\ulcorner\varphi\urcorner) \iff \varphi(\ulcorner\varphi\urcorner) \text{ is provable,}$$

that is,

$$\chi^*\,R\,\varphi \iff \varphi\,Q\,\varphi,$$

and hence

$$(\neg\chi^*)\,R\,\varphi \iff \neg(\varphi\,Q\,\varphi).$$

By Theorem 2, the sentence $\neg\chi^*(\ulcorner\neg\chi^*\urcorner)$ is true, but not provable. Thus the notion of provability is incomplete in the sense of not establishing all truths.

4 Criteriological science

In Theorem 1, let R be the relation of studying, or examining, or criticizing, or giving an account of. This is the relation that a science has to its objects. But some things have this relation to themselves. Thinking is an example: thinking is self-critical. By the theorem, there is no science whose objects are precisely those things that are not self-critical. Any science that studies all non-self-critical things must also study self-critical things. Such a science must therefore be *criteriological*: it must give an account of how this self-criticism is done.

Collingwood defines criteriological science in *An Essay on Metaphysics* of 1940 [3, pp. 108–9]. He mentions the traditional criteriological sciences of logic and ethics: the sciences of theoretical and practical thought, respectively. Collingwood prefers the term *criteriological* over the more usual term *normative,* since the latter suggests a science that applies its own criteria or norms, rather than accounting for the criteria used by what it studies. If I don't like what you do, it is not scientific just to declare that what you do violates my own ethical criteria; I should try to understand whether it violates your own criteria, and how you arrive at these criteria.

By speaking about what I *should* do as an ethicist, I demonstrate that the science of ethics applies also to itself. Collingwood uses such an argument in his 1933 *Essay on Philosophical Method* [4, pp. 131-3] to prove that philosophy is "categorical" in the sense that "its subject-matter is no mere hypothesis, but something actually existing" [4, p. 127]. Ethics and logic are philosophical sciences.

Thus Collingwood gives an "ontological proof" in the *Essay* of 1933. Mathematics works with hypothetical objects; but philosophy must conceive of its objects as actually existing. Collingwood makes it clearer in the *Essay* of 1940 that the ontological proof of the existence of God is really a proof of belief, a proof that God is an "absolute presupposition" of anybody who believes in the unity of the world. Thales is such a person: his argument that all is water is an introduction of monotheism to a polytheistic world.

Natural sciences are not criteriological. Psychology is supposed to be a natural science of thought. But there can be no such science of thought as such.

Collingwood does not use the term "criteriological" in the earlier *Essay.* He does use it in the 1938 *Principles of Art* [1, p. 171 n.]. He argues in the 1939 *Autobiography* [2, pp. 94–5] that psychology cannot be a science of mind as such, because psychology is not criteriological; but he does not actually use this term.

References

[1] R. G. Collingwood. *The Principles of Art.* Oxford University Press, London, Oxford, and New York, paperback edition, 1958. First published 1938.

[2] R. G. Collingwood. *An Autobiography.* Clarendon Press, Oxford, 1978. First published 1939. With a new introduction by Stephen Toulmin. Reprinted 2002.

[3] R. G. Collingwood. *An Essay on Metaphysics.* Clarendon Press, Oxford, revised edition, 1998. With an Introduction and additional material edited by Rex Martin. Published in paperback 2002. First edition 1940.

[4] R. G. Collingwood. *An Essay on Philosophical Method.* Clarendon Press, Oxford, new edition, 2005. With an Introduction and additional material edited by James Connelly and Giuseppina D'Oro. First edition 1933.

[5] Kurt Gödel. On formally undecidable propositions of *principia mathematica* and related systems I. In van Heijenoort [8], pages 596–616. First published 1931.

[6] Bertrand Russell. Letter to Frege. In van Heijenoort [8], pages 124–5. First published 1902.

[7] Alfred Tarski. The concept of truth in formalized languages (1933). In *Logic, semantics, metamathematics*, pages 152–278. Hackett Publishing Co., Indianapolis, IN, second edition, 1983. Papers from 1923 to 1938, Translated by J. H. Woodger, Edited and with an introduction by John Corcoran.

[8] Jean van Heijenoort, editor. *From Frege to Gödel: A source book in mathematical logic, 1879–1931.* Harvard University Press, Cambridge, MA, 2002.