Anselm's Argument and Berry's Paradox[*]

P. Schlenker

(UCLA & Institut Jean-Nicod)

Abstract: We argue that Anselm's ontological argument (or at least one reconstruction of it) is based on an empirical version of Berry's paradox. It is invalid, but it takes some understanding of trivalence to see why this is so. Under our analysis, Anselm's use of the notion of existence is *not* the heart of the matter; rather, trivalence is.

Anselm's 'proof' of the existence of God goes like this: I have the concept of *that than which nothing greater can be thought*. If this concept were not instantiated, something greater could be thought - namely the same concept with the property of existence. This is a contradiction, therefore *that than which nothing greater can be thought* is instantiated - God exists. We suggest that this argument is based on an empirical version of Berry's paradox. It is invalid, but it takes some understanding of trivalence to see why this is so. Under our analysis, Anselm's use of the notion of existence is *not* the heart of the matter; rather, trivalence is (as far as we know, this distinguishes the present reconstruction from its predecessors, notably Lewis 1970; see also Oppy 2007 for references).

For the sake of clarity, we start by explaining with some simple examples why someone who always reasoned within a bivalent framework would make systematic errors when faced with non-trivial empirical paradoxes. In particular, we develop an empirical version of Berry's paradox which is similar in structure to Anselm's argument. With these tools in place, we offer a simple reconstruction of Anselm's argument, one in which the notion of existence is entirely innocuous.

1 Paradoxes and Trivalence

Four important lessons can be drawn from the recent study of semantic paradoxes (Kripke 1975):

(i) A semantics for a paradoxical object language should be (at least) trivalent, so that paradoxical statements can be given an indeterminate truth value (neither true nor false).

(ii) Paradoxes may arise in devious ways. The Liar (*This very sentence is not true)* is particularly simple because it involves direct self-reference and a truth predicate. But paradoxes may be obtained with a denotation or a satisfaction predicate instead (McGee 1992), and direct self-reference may be replaced with circularity (= indirect self-reference), or with self-referential quantification (paradoxes without circularity can also be obtained when infinite series of sentences are considered, as in Yablo 1993[1]).

(iii) Some statements may display a pathological behavior even though they may sometimes obtain a classical truth value. This is in particular the case of empirical Liars, which only display a pathological behavior when certain empirical conditions hold. A particularly simple example is provided by a sentence named ε, which says: *It is raining or ε is not true*. When it is raining, the first disjunct is true, and hence the sentence is too (on the assumption, which we will make throughout, that sentences are evaluated within Kleene's strong trivalent logic);

[*] For discussion and comments, I wish to thank Paul Egré, Béatrice Lenoir, and an audience at École Normale Supérieure (Lyon). The author gratefully acknowledges the financial support of the American Council of Learned Societies ('Ryskamp Fellowship') and of UCLA.

[1] In fact, it follows from results reported in Schlenker 2007 that under broad conditions *every semantic phenomenon - and hence every paradox - obtained with self-reference can be emulated without it.*

when it is not raining, the sentence is paradoxical. More complex examples will be constructed below.
(iv) Finally, formal semantic accounts of paradoxes are typically plagued by *revenge problems*: strengthened forms of classical paradoxes must typically be excluded from the object language on pain of making the logic inconsistent. Such is the case of the Strengthened Liar: *This very sentence is something other than true* cannot be true or false, for the familiar reasons; but it also can't be indeterminate, because if so it would be *something other than true* - and hence it should be true.

Now suppose that a philosopher who was innocent of these developments was faced with a particularly devious empirical paradox, one that was always true or indeterminate (for instance the sentence ε defined above). What would he think? He would reason bivalently, and he would overlook the possibility that the sentence may be indeterminate. From the observation that the sentence is never false, he would jump to the conclusion that it is *a priori* true. We will suggest that this is the heart of Anselm's error, and that his argument is an empirical version of Berry's paradox - to which we now turn.

2 Berry's Paradox: Simple and Empirical Versions

We henceforth assume Kripke's theory of truth. Standard results (Kripke 1975, McGee 1992) guarantee that interpretations can be found in which *true* is true of the true sentences, false of the non-sentences and of the false sentences, and indeterminate of the indeterminate sentences; similar results apply to the binary predicate *is true of*.

2.1 A Simple Version of Berry's Paradox

Berry's paradox can be stated informally as follows: there are only finitely many integers that can be defined in fewer than thirty syllables; therefore there exists an integer which is the *smallest integer not definable in fewer than thirty syllables*. But the italicized description defines this integer, and it comprises fewer than thirty syllables - contradiction.

To make things precise, we provide in (1) a semi-formal rendering of the paradox. We will say that a formula *F* with one free variable defines o just in case o is the one and only object that makes *F* true. Since the number of symbols needed to define our formula will depend on the details of the formal language under consideration, we replace *thirty syllables* with *n symbols*, leaving it to each implementation to pick a value for n which is large enough to make the paradox go through[2]:

(1) x is an integer $\wedge$ $\neg\exists$y (y is a formula true of exactly one object[3] $\wedge$ y contains fewer than n symbols $\wedge$ y is true of x) $\wedge$ $\forall$x' ([x' is an integer $\wedge$ $\neg\exists$y'(y' is a formula true of exactly one object $\wedge$ y' contains fewer than n symbols $\wedge$ y' is true of x'] $\Rightarrow$ x' $\geq$ x)

In words, Berry's formula - henceforth called $\beta[x]$ - says that *x is an integer not definable in fewer than n symbols, and every integer not definable in fewer than n symbols is at least as great as x*. With a large enough value for n, Berry's formula is paradoxical, in the sense that it is indeterminate of at least some integers. For if not, we could reason as follows:
-Suppose that $\beta[x]$ is true of some integer i. Then it can *only* be true of i (if it is true of i', then both i and i' satisfy the second conjunct, and by taking into account the third conjunct we can conclude that i $\geq$ i' and that i' $\leq$ i - hence i = i'). Since by assumption $\beta[x]$ always has a

[2] This assumes that the object language has a reasonably compact way of referring to some large enough integers (or else the formula in (1) might contain more than n symbols for *any* value of *n*).
[3] *y is a formula true of exactly one object* abbreviates:
(i) $\exists$x'' (y is a formula true of x'' $\wedge$ $\forall$x''' (y is true of x''' $\Leftrightarrow$ x''' = x''))

classical value, it is false of all other integers - and thus it defines i. On the assumption that n is large enough, this contradicts the assumption that i cannot be defined in fewer than n symbols.
-On the other hand *β[x]* couldn't be false of *all* integers. For if it were, we could reason as follows: only finitely many integers can be defined in fewer than n symbols, and thus only finitely many integers make the second conjunct false. Any other integer i makes it true or indeterminate. In order for i to make *β[x]* false, i must make the third conjunct false. This implies that we can find an integer u not definable in fewer than n symbols which is strictly smaller than i. But u makes the second conjunct true; it must thus make the third conjunct false, which means that there is a number u' not definable in less than n symbols which is strictly smaller than u. By iterating the reasoning, we could show that 0 is not definable in fewer than n symbols - which cannot be (it is definable by the formula *(x = 0)*).

Of course we could try to define a *bivalent* predicate *is true* of* which is true of a pair <F, d> just in case *F* is true of d, and which is false of <F, d> in all other cases. This would allow us to construct a formula *β*[x]* identical to *β[x]*, with *is true* of* replacing *is true of*. But then we could reproduce with *β*[x]* the reasoning that just showed that *β[x]* can't be classical - hence a contradiction. This should come as no surprise: *is true* of* is the satisfaction predicate corresponding to a modified truth predicate *true** which is true of the true sentences and false of everything else. But it is immediate that *true** makes it possible to define a Strengthened Liar (e.g. *This sentence is not true**). There is no completely satisfactory solution to revenge paradoxes, and we will see that the same weakness extends to the analysis of Anselm's argument.

2.2 An Empirical Version of Berry's Paradox

Let us now call 'concrete number' any integer that has been written down in the history of the universe. There are possible worlds in which there is a greatest concrete number, and others in which there is none, either because the set of concrete numbers is empty or because it has no upper bound. Clearly, there are only finitely many concrete numbers that can be described in fewer than n symbols. We can thus talk of the *greatest concrete number that can be described in fewer than n symbols*, a formula defined in (2):

(2) x is a concrete number ∧ ∃y (y is a formula true of exactly one object ∧ y contains fewer than n symbols ∧ y is true of x) ∧ ∀x' ([x' is a concrete number ∧ ∃y' (y' is a formula true of exactly one object ∧ y' contains fewer than n symbols ∧ y' is true of x'] ⇒ x' ≤ x)

In words, this formula - call it *β'[x]* - says that *x is a concrete number that can be described in fewer than n symbols, and every concrete number that can be described in fewer than n symbols is at most as great as x.* There are three main cases to consider to determine what *β'[x]* is true of:
1. If there are no concrete numbers, *β'[x]* is true of nothing at all (because its first conjunct is always false).
2. If there is a greatest concrete number, which we call *i*, *β'[x]* is true of i and nothing else. This is because i is defined by the (classical) formula in (3):

(3) x is a concrete number ∧ ∀x' (x' is a concrete number ⇒ x' ≤ x)

If n is large enough, (3) defines i and guarantees that the second conjunct of *β'[x]* is true of i. Furthermore, when *x* denotes i the third conjunct of *β'[x]* is true because for every x', either x' > x, and hence x' is not concrete, in which case the antecedent of the material implication is false, and thus the implication itself is true; or x' ≤ x, in which case the implication is true

because its consequent is. It is also clear that the third conjunct is not true of any other concrete number. In sum, *β'[x]* defines i.
3. If there is at least one concrete number but no greatest one, and if *n* is large enough, *β'[x] is indeterminate of some integers and false of all others.* As in the analysis of the standard version of Berry's paradox, we assume, for contradiction, that *β'[x]* is always true or false.
3.1. Suppose that *β'[x]* is true of some integer i. Then it can *only* be true of i (if it were true of some integer i', we would have that i and i' both satisfy the second conjunct, and by taking into account the third conjunct we could conclude both that i ≥ i' and that i' ≤ i). Since by assumption *β'[x]* always has a classical value, it is false of all other integers - and thus it defines i. In turn, we can define a formula *β"[x]* as in (4), where *x follows z* abbreviates *x is a concrete number* $\wedge\ \forall x'\ ((x' > z\ \wedge x'$ *is a concrete number*$) \Rightarrow x' \geq x)$:

(4) $\exists z$ (x follows z $\wedge$ β'[z/x])

(4) contains fewer than n symbols, and the second conjunct of (4) is true of i and nothing else. It follows that this formula itself is true of the follower of i and nothing else - which contradicts the hypothesis that i satisfies *β'[x].*
3.2. On the other hand *β'[x]* couldn't be false of *all* integers. Consider the concrete integers that are definable in fewer than n symbols. One of them, call it *i*, is the greatest such number. Clearly, i satisfies the first two conjuncts of *β'[x].* Thus it must make the third conjunct false; but this entails that there is a concrete number definable in fewer than n symbols which is greater than i - *contra* hypothesis.

On the assumption that there *are* concrete numbers, a bivalent philosopher could 'prove' that there is a greatest one by arguing like this: there are only finitely many concrete numbers definable in fewer than n symbols, and thus one of them must be the greatest such number - it must make the formula *β'[x]* in (2) true. But the philosopher's argument is incorrect: in Case 3, *there is simply no number that makes the formula true*; all numbers make it false or indeterminate. As in the simple version of Berry's paradox, however, the trivalent solution suffers from a revenge problem: as soon as a predicate *is true* of* is introduced in the language, Case 3 becomes intractable again because the formula *β'*[x]* (where *true** replaces *true*) is bivalent by construction.

3 Anselm's Argument

Let us turn to Anselm's argument, which we claim to be similar to the empirical version of Berry's paradox. Talk of describable concrete numbers will be replaced with talk of concept-like things, with the assumption that concepts are ordered, and that a concept that is not instantiated has infinitely many followers. The modified form of the argument will 'prove' (erroneously) that there exists a greatest concept, and thus that it must be instantiated.

Anselm states his argument as follows:

> So even the Fool is convinced that something than which nothing greater can be thought is at least in his understanding; for when he hears of this [being], he understands [what he hears], and whatever is understood is in the understanding. But surely that than which a greater cannot be thought cannot be only in the understanding. For if it were only in the understanding, it could be thought to exist also in reality—something which is greater [than existing only in the understanding]. (...) But surely this [conclusion] is impossible. Hence, without doubt, something than which a greater cannot be thought exists both in the understanding and in reality. [translation: Hopkins and Richardson]

We propose to paraphrase the argument as follows: (i) One has the concept of *something than which nothing greater can be thought*. (ii) Suppose, for contradiction, that the concept of *that than which nothing greater can be thought* was not instantiated, i.e. that nothing fell under this concept in the world as we know it. Then something greater could be thought, namely the same concept *with* existence. (iii) Therefore *that than which nothing greater can be thought* is instantiated.

Our paraphrase cuts some logical corners, in particular in the transition from an indefinite description in (i) (*something than which...*) to a definite description in (ii) (... *that than which* ...). But we will now make the reasoning more explicit and conclude that the only essential error lies in the fact that the argument is developed within a bivalent framework.

1. Domain

-Since we do not wish to make any controversial assumptions about the treatment of existence, we consider a domain of objects that includes concepts (which may for instance be thought of as functions from possible worlds to individuals) and also pairs of the form <c, ∃>, where c is a concept. ∃ is intended to be the property of existence, but *nothing in its interpretation will matter besides the conditions we give below* (this is the sense in which our treatment of existence is entirely innocuous). Formally, we start with the definition of the proto-objects derived from a concept c:

• If c is a concept, c is a proto-object derived from c.

• If o is a proto-object derived from c, so is <o, ∃>.

o is a proto-object just in case for some concept c, o is a proto-object derived from c.

-The argument posits that if a concept c is not instantiated, c with existence, i.e. in our words <c, ∃>, is greater than c. But the argument does not tell us what happens when a concept c *is* instantiated. It is plausible that in this case <c, ∃> should be taken to be *the same thing* as c. Accordingly, we define an equivalence relation $=^e$ by: a $=^e$ b iff for some concept c, c is instantiated, and both a and b are proto-objects derived from c. If o is a proto-object, we write the equivalence class of o under $=^e$ as *[o]*.

-The domain of objects is the set of equivalence classes of $=^e$ over the proto-objects.

2. Ordering

We assume that a relation < can be defined, which ensures:

(i) that for any proto-object o derived from c, if c is not instantiated (in the world as we know it), then [o] < [<o, ∃>].

(ii) that for any object o of the domain, either

(a) o is the maximum of <, or

(b) o has a successor according to <, i.e. there is an object o' such that o < o', and no o" is such that o < o" < o'.

Anselm certainly had in mind some 'natural' interpretation of <. We do not wish to take a stand on this issue, but we note that it is straightforward to *define* such an ordering when the concepts form a set[4]. On the other hand a non-trivial consequence of our analysis should be highlighted: if a concept c is not instantiated, then [c] < [<c, ∃>], as desired; but it is also the case that [<c, ∃>] < [<<c, ∃>, ∃>] < [<<<c, ∃>, ∃>, ∃>], etc. This follows because <c, ∃>, <<c, ∃>, ∃>, etc. are all proto-objects derived from c; since c is not instantiated, (i) applies, and leads to this infinite hierarchy. This will be crucial below.

We can now reconstruct Anselm's argument.

[4] By the axiom of choice, the set of concepts can be well-ordered by a certain relation **<**. We extend **<** to all objects through a definition by cases:

-If c is instantiated, for any proto-object o derived from c, [c] = [o], so nothing needs to be added to the extension of **<**.

-If c is not instantiated, and if o and o' are proto-objects derived from c, we say that [o'] is the *mother* of [o] just in case o' = <o, ∃>. We take the *ancestor* relation to be the transitive closure of the *mother* relation, and we define: [o] **<** [o'] just in case [o'] is an ancestor of [o].

1. Premise: Being a finite creature, I can only produce finitely many formulas, hence finitely many descriptions.

2. Consequence [incorrect]: Since < is a complete ordering, something - call it G - makes the following formula true:

(5) x is the greatest object I can define

3. Reduction: Suppose, for contradiction, that G is not the maximum of <. By the definition of <, G has a successor. But certainly I can produce the following formula:

(6) x is a successor according to < of the greatest object I can define

(6) defines the successor of G. But this is a description I just produced, so G is not the greatest object I can define, *contra* hypothesis.

4. Conclusion: Therefore G is the maximum of <. For some proto-object o, G = [o]. If o is a proto-object derived from a concept which is not instantiated, <G, ∃> > G, which contradicts the assumption that G is the maximum of >. Therefore o is derived from a concept c which *is* instantiated, and thus G = [c].

As noted, whenever a proto-object o is derived from a concept c that is not instantiated, we obtain an infinite hierarchy [c] < [<c, ∃>] < [<c, <c, ∃>>], etc; by contrast, when c *is* instantiated, any proto-object o derived from c guarantees that [o] = [c]. Thus *if there exists a greatest object, it must be derived from a concept that is instantiated* - which is crucial to obtain the desired conclusion in Step 4. We *could* have defined things differently; we could have stipulated that whenever c is not instantiated, [c] < [<c, ∃>], but [<c, ∃>] = [<<c, ∃>, ∃>] = [<<<c, ∃>, ∃>, ∃>], etc. In that case, Step 4 would have produced a much weaker conclusion, namely that either (a) G = [c], where c is a concept which is instantiated, or (b) G = [<c, ∃>] (> [c]), where c is a concept that is *not* instantiated. But the second possibility is entirely devoid of theological interest, since it certainly does not show that God exists. By contrast, with the assumptions that we did make, the conclusion yields the desired result. However, the reasoning is invalid. Let us see why.

I take the premise to be unproblematic if *I can only produce finitely many formulas* is understood to mean that *throughout my lifetime I will only produce finitely many formulas*, or alternatively that *my brain has only finitely many states, which can only produce finitely many formulas*. What is not meant, of course, is the absurd claim that English (or Latin, for that matter) can only produce finitely many formulas.

The crucial step is the consequence drawn in Step 2. To understand what is wrong with it, we must give a somewhat more formal definition of Anselm's description, which we will call $\alpha[x]$:

(7) x is an object ∧ ∃y (y is a formula true of exactly one object ∧ I can think y ∧ y is true of x) ∧ ∀x' [∃y' (y' is a formula true of exactly one object ∧ I can think y' ∧ y' is true of x] ⇒ x' ≤ x)

We then reason as in our analysis of the empirical version of Berry's paradox stated in (2):
1. It is clear that there are objects, so the first case we discussed after (2) does not arise here.
2. If there is a greatest object [G], G must be derived from a concept g which is instantiated, and thus [G] = [g]. Furthermore, [G] can be defined by the classical formula in (8), which is analogous to the formula that defined the 'greatest concrete number' in (3):

(8) x is an object ∧ ∀x' (x' is an object ⇒ x' ≤ x)

When x denotes [G], the first conjunct of $\alpha[x]$ (i.e. (7)) is trivially true, the second is true due to the existence of (8), and the third conjunct is true because x denotes the greatest object; in addition, the third conjunct is true of no other object. Therefore $\alpha[x]$ defines [G].

3. If there is no greatest object, we can show as in our discussion of the empirical version of Berry's paradox that $\alpha[x]$ must be indeterminate of some objects and false of all others (since the assumption that it is true leads to a contradiction).

As was the case with the empirical version of Berry's paradox, someone who reasoned bivalently would think that some object must make $\alpha[x]$ true, and hence be defined by it. Since a contradiction arises in Case 3, we must be in Case 2. But once we reason within a trivalent framework, we see that no object makes $\alpha[x]$ true in Case 3: all objects make the formula false or indeterminate, and no contradiction arises. Anselm's error is to reason bivalently in a language which is trivalent.

One final point is that the trivalent analysis comes with a price: we must banish from the object language the bivalent predicate *is true* of,* on pain of making Case 3 intractable again. This is just as it should be: in our analysis, Anselm's argument is just an empirical version of Berry's paradox. The latter has a strengthened version, and so does the former. In the end, Anselm's argument is just as hard to treat as the semantic paradoxes.

References

Anselm of Canterbury: 1077-1078?, *Proslogion*.

Boolos, G.: 1989, A New Proof of the Gödel Incompleteness Theorem. *Notices of the American Mathematical Society* 36: 388-390

Boolos, G., Burgess, J. and Jeffrey, R.: 2002, *Computability and Logic* (fourth edition). Cambridge University Press.

Chaitin, G.: 1995, The Berry Paradox. *Complexity* 1, 1: 26-30.

Kripke, S.: 1975, Outline of a Theory of Truth. *Journal of Philosophy* 72: 690–716.

Lewis, D.: 1970, Anselm and Actuality. *Noûs* 4, 175-188

McGee, V.: 1992, *Truth, Vagueness, and Paradox.* Hackett.

Oppy, G.: 2007, Ontological Arguments. *Stanford Encyclopedia of Philosophy*. Available online at http://plato.stanford.edu/entries/ontological-arguments/

Schlenker, P.: 2007, The Elimination of Self-Reference (Generalized Yablo-Series and the Theory of Truth). *Journal of Philosophical Logic* 36, 3: 251-307.

Uzquiano, G.: 2004, An Infinitary Paradox of Denotation. *Analysis* 64.2: 128–31.

Yablo, S.: 1993, Paradox without self-reference. *Analysis* 53: 251–52.