Arhat Virdi

The slingshot argument, Godel’s hesitation and Tarskian semantics

Article (Accepted version)
(Refereed)

The slingshot argument, Gödel's hesitation and Tarskian semantics

ARHAT VIRDI

The slingshot argument is a reductio purporting to show that if there are facts at all there is only one to which all true statements correspond. If facts are not non-trivially individuable then this presumably must render the notion of fact and, by implication, theories such as the correspondence theory of truth incoherent. Church and Davidson (among others) deployed the slingshot in exoneration of the Fregean conclusion that there is a uni-referent – the 'True' – for all true statements. The slingshot relies crucially on treating definite descriptions as singular, referring terms, a treatment that is rendered unnecessary on Russell's theory of descriptions. If this is so, friends of facts such as Russell can rest content. I, however, argue against the thesis that Russell's theory so succeeds and develop what Gödel could have meant when, in thinking about this application of Russellian semantics, was prompted to write: "I cannot help feeling that the problem raised by Frege's puzzling conclusion has only been evaded by Russell's theory of descriptions and that there is something behind it which is not yet completely understood." (1944: 215). I conclude by suggesting that the coarse-grained, folk theory of facts to which the slingshot objection incontestably applies is in need of being fine-grained into a scientifically more sophisticated theory, and that such an account is to be found in a Tarskian definition of truth which, moreover, also succeeds in placing the correspondence theory of truth on a secure and satisfactory footing.[1]

1. *The slingshot argument and Russellian semantics*

The slingshot was first developed by Alonzo Church. In *Introduction to Mathematical Logic* (1956: 24-25) he considers the following set of sentences:

(1) Sir Walter Scott is the author of *Waverley*
(2) Sir Walter Scott is the man who wrote twenty-nine *Waverley* novels altogether
(3) The number, such that Sir Walter Scott is the man who wrote that many *Waverley* novels altogether, is twenty-nine
(4) The number of counties in Utah is twenty-nine

Given that the name or description 'the author of *Waverley*' is replaced by another ('the man who wrote twenty-nine *Waverley* novels altogether') which has the same reference, i.e. Scott, (1) and (2) must have the same reference. The same applies to (3) and (4): the latter is obtained from the former by replacing the description 'the number, such that Sir Walter Scott is the man who wrote that many *Waverley* novels altogether' by another referring to the same object (the number twenty-nine). Given that (2) and (3) are "if not synonymous...[then] at least so nearly so as to ensure its having the same denotation" for Church, they too must have the same reference. Therefore, (1) and (2), (2) and (3), and (3) and (4) have the same reference when taken pairwise, which means that (1) and (4) must do too. So, (1) and (4) have differing senses *yet* the same reference. The only semantic feature that they retain is their truth-value. Church used this example to demonstrate that sentences with non-equivalent senses may still have equivalent referents, and that no matter what the reference is each will have the same one.

Davidson lost no time in using this argument explicitly against the correspondence theory of truth. In "True to the Facts", Davidson (1969: 41) first considers when

[1] I thank an anonymous referee for useful suggestions on improving the original draft.

(S) The statement that *p* corresponds to the fact that *q*

would be true. Clearly **(S)** is true when both *p* and *q* are replaced by the same sentence. However, unless facts are to be understood as mere reflections of true sentences, there ought to be true instances of **(S)** where *p* and *q* are not identical. Davidson then observes that since (as an example) Naples satisfies the following description 'the largest city within thirty miles of Ischia', then the statement that Naples is farther north than Red Bluff corresponds to the fact that Red Bluff is farther south than the largest city within thirty miles of Ischia. Given further that Naples *also* satisfies the description 'the largest city within thirty miles of Ischia, and such that London is in England', then "we begin to suspect that if a statement corresponds to one fact, it corresponds to all." (1969: 42). This suspicion is validated as long as the following two principles are assumed to hold:

- The statements replacing '*p*' and '*q*' are logically equivalent
- '*p*' and '*q*' differ only in that a singular term has been replaced by a co-extensive singular term

For Davidson then, the argument is this:

> Let '*s*' abbreviate some true sentence. Then surely the statement that *s* corresponds to the fact that *s*. But we may substitute for the second '*s*' the logically equivalent '(the *x* such that *x* is identical with Diogenes and *s*) is identical with (the *x* such that *x* is identical with Diogenes)'. Applying the principle that we may substitute coextensive singular terms, we can substitute '*t*' for '*s*' in the last quoted sentence, provided '*t*' is true. Finally, reversing the first step we conclude that the statement that *s* corresponds to the fact that *t*, where '*s*' and '*t*' are any true sentences. (Davidson 1969: 42)

Formally, the argument looks like this {where '(ι*x*)' means 'the *x* such that…*x*…'}:

1. s	Premise
2. $(\iota x)(x = d \wedge s) = (\iota x)(x = d)$	From 1., given substitution *salva veritate* of logical equivalents
3. $(\iota x)(x = d \wedge t) = (\iota x)(x = d)$	From 2., given substitution *salva veritate* of co-referring terms
4. t	From 3., given substitution *salva veritate* of logical equivalents

All four lines of this argument correspond to the same fact. In "The Structure and Content of Truth", Davidson argued that the moral to draw from this is that it:

> …trivialize[s] the concept of correspondence completely; there is no interest in the relation of correspondence if there is only one thing to which to correspond, since, as in any such case, the relation may well be collapsed into a simple property: thus, "*s* corresponds to the universe", like "*s* corresponds to (or names) the True", or "*s* corresponds to the facts" can less misleadingly be read "*s* is true". (Davidson 1990: 303)

In fact, as Gödel indicated in "Russell's Mathematical Logic" (1944: 213-214), the slingshot being loaded here can be made even more powerful. Gödel employed a notion of equivalence weaker than that of logical equivalence – what Stephen Neale has termed *Gödelian equivalence*[2] – one obtaining between sentences like 'Fa' and '$a = (\iota x)((x = a) \wedge Fx)$':[3] they are to "mean the same thing". This is weaker because true identity statements remaining true under the substitution of logically equivalent statements entails *Gödelian equivalence*, but not vice-versa. The argument runs through assuming this weaker equivalence principle.

There remains a question, however. The validity of the slingshot argument presumably depends on the validity of the semantics of definite descriptions adopted. Such a semantics, it is argued, must – in the case of the Church-Davidson version – (i) render 's' and '$\iota x(x = d \wedge s) = \iota x(x = d)$' logically equivalent, (ii) declare the definite descriptions '$\iota x(x = d \wedge s)$' and '$\iota x(x = d \wedge t)$' co-referential when 's' and 't' are true, and (iii) treat definite descriptions as singular, referring terms. And similarly for the Gödelian version: if one wished to hold that definite descriptions are singular terms that refer and that sentences standing for facts are determined by the referents of their component parts then one cannot hold that 'Fa' is somehow a different fact from the fact that '$a = (\iota x)((x = a) \wedge Fx)$'. However, as many have pointed out, on Russell's theory of descriptions,[4] definite descriptions do not stand for objects, or refer to things; they are not referential because they are not singular terms. According to this theory, any sentence of the form 'the F is G' ought rather to be understood as belonging to the class of quantificational-predicational expressions; they are on a par with the quantifiers 'every', 'some', 'a', 'no' which are syncategorematic terms that, when combined with nominal expressions, yield noun phrases (*cf*. Russell 1905: 42). Thus, the sentence 'the F is G' is equivalent to the corresponding sentence 'there is one and only one F, and it is G', formalized as '$\exists x(Fx \wedge \forall y(Fy \rightarrow y = x) \wedge Gx)$', giving us the wherewithal to turn any sentence containing definite descriptions into an equivalent that is description-free. Given this theory, and the 'principle of compositionality', it cannot then be the case that both '$a = (\iota x)((x = a) \wedge Fx)$' and '$a = (\iota x)((x = a) \wedge a \neq b)$' are obtainable from each other from the substitution of co-referring terms. And so, it does not follow that they stand for the same fact. The property of being F is part of the fact corresponding to '$a = (\iota x)((x = a) \wedge Fx)$' but not the fact corresponding to '$a = (\iota x)((x = a) \wedge a \neq b)$'. Indeed, '$Fa$' has a truthmaker that is an entirely different (singular) fact from the general fact making '$a = (\iota x)((x = a) \wedge Fx)$' true, and so a Russellian need not accept that they stand for the same fact. On Russell's theory, '$a = (\iota x)((x = a) \wedge Fx)$' is shorthand for '$\exists y[((y = a) \wedge Fy) \wedge \forall z(((z = a) \wedge Fz) \rightarrow (z = y)) \wedge (a = y)]$'[5] and so its truthmaker – a – need not be the truthmaker of 'Fa'.

2. *Gödel's hesitation*

Gödel, however, was hesitant to endorse this application of Russell's semantics, saying:

> ...I cannot help feeling that the problem raised by Frege's puzzling conclusion [that all true sentences have the same signification] has only been evaded by Russell's theory of descriptions and that there is something behind it which is not yet completely understood. (Gödel 1944: 215)

---

[2] See Neale 1995, 1997 & 2001.

[3] Gödel also assumed that any sentence standing for a fact can be rephrased into predicate-argument form (*cf*. Gödel 1944: 214, footnote 5). Clearly, without this assumption the slingshot envisaged would only hold for all true *atomic* sentences.

[4] See Russell 1905, 1918.

[5] Strictly speaking, Russell's theory that definite descriptions are 'incomplete' means that they have to analyzed *within a sentence*; they are not themselves sentences (nor equivalent to sentences).

Gödel was right to be hesitant. Excluding definite descriptions from the primitive notation just creates the illusion of a solution, since, as Church (1943) showed, the argument can be reformulated in terms of set-abstraction operators where there is no question that they refer (in the standard model, whose existence we can reasonably assume here). This is what Davidson's slingshot looks like when the *iota* operators are replaced by set abstracts: let *s* and *t* abbreviate true sentences. The following then is a *valid* argument, with each line corresponding to the same fact:

1. s	Premise
2. $\{x: x = d \wedge s\} = \{x: x = d\}$	From 1., given substitution *salva veritate* of logical equivalents
3. $\{x: x = d \wedge t\} = \{x: x = d\}$	From 2., given substitution *salva veritate* of co-referring terms
4. t	From 3., given substitution *salva veritate* of logical equivalents

To argue, as Neale does, that Gödel's slingshot "forces philosophers to say something about the semantics of definite descriptions…as soon as they posit entities to which sentences are meant to correspond" (2001: 223) would, therefore, be wrong-headed: too much weight has been placed on imaginary problems concerning the *iota* operator.[6] Does the slingshot argument force us to revise our ordinary speech permitting us to speak, as it appears to be doing, of nothing more than one fact? What the argument does demonstrate is that the folk theory of facts has (quite possibly) unacceptable consequences, and rather graphically exhibits one. It shows that there is a need to move from folk, fact-based semantics, which doesn't work properly, to a more scientific semantics. The conceptual apparatus provided by Tarski, it will be demonstrated, succeeds where the 'folk' fact-talk failed: facts are a very important, almost indispensable, ontological category whose intensional structure is actually preserved in a theory like Tarski's and it is owing to this kind of faithfulness to the structure of facts which serves to deflect the slingshot.

3. *Tarskian semantics*

Fortunately, the conceptual apparatus provided by Alfred Tarski succeeds where the 'folk' fact-talk failed, and collaterally the account of facts provided in a Tarskian truth definition (*pace* Davidson et al) allows us to see precisely how true sentences correspond to facts: true sentences are homomorphic images of facts, i.e. a true sentence represents, in a form-preserving manner, the truth-making facts in it. To see all this, we need to recount the clauses of a Tarskian truth definition. In his mature work, Tarski adopted the convention according to which the non-logical constants of a language L are enumerated in a fixed order and their interpretations in a relational structure are then given in the same order. So, for a given interpretation function I from L to a domain of individuals X, we can treat (L, I) as an interpreted language and define a structure for this language as a *relational system* of the form (X, s_1,…, A_1,…, R_1,…) with designated elements s, subsets A and relations R that are the I-images of the vocabulary of L in the domain X. Tarski did not make the interpretation function I explicit but it is clear that such a function from L to the relational system is presupposed.[7] In textbooks on model theory, it has become standard to explicitly express the link

---

[6] This does not mean that the friend of facts has nothing instructive to say; only that she is not forced into making any commitments about the *semantics* of descriptions (*cf*. Rodriguez-Pereya 2003).

[7] This is arguably confirmed by the fact that in his *Introduction to Semantics*, Carnap's own characterization of Tarski's approach (of which he was a great admirer) a *designation* function Des corresponding to the pair (L, I) is explicitly given in Carnap's semantical system S. For Carnap, Des is a language-world relation where individual constants designate individual objects and predicates designate properties. Des is first defined for individuals and predicates and then by recursion for sentences. For example, if *a* designates snow {DesInd('*a*', snow)}and *P* designates the property of being white {DesAttr('*P*', the property of being white)}, then *P*(*a*) designates the proposition that snow is white {DesProp('*P*(*a*)', snow is white)}. Truth of sentences in semantical system S is defined as follows:

between language L and a set-theoretic structure **M** by the interpretation function I.[8] An L-structure **M** is thus defined as the pair (X, I). Language L is then an uninterpreted syntactic language which becomes interpreted via **M**. We follow the standard modern formulation.[9] For atomic sentences of L, truth in structure $\mathbf{M}=(X, \mathrm{I})$ is defined by the following conditions:

$$\mathbf{M}\surd P(a_1) \text{ iff } I(a_1)\in I(P)$$
$$\mathbf{M}\surd Q(a_1,a_2) \text{ iff } <I(a_1), I(a_2)>\in I(Q)$$

An open formula A of L with free variable x_i is assigned a truth value in the structure **M** by some element of X. Let $s = (s_1, s_2, s_3,\ldots)$ be an infinite sequence of objects from X. Then sequence s satisfies formula A in structure **M**, i.e. the relation $\mathbf{M}\surd_s A$, is defined by recursion on the complexity of A. For example:

$$\mathbf{M}\surd_s A\vee B \text{ iff } \mathbf{M}\surd_s A \text{ or } \mathbf{M}\surd_s B$$
$$\mathbf{M}\surd_s \forall x_i A \text{ iff } \mathbf{M}\surd_{s(i/b)} A \text{ for all } b\in X$$

where s(i/b) is the sequence obtained from s by replacing the i-th element of s with b. The basic clauses for atomic formulas take the following form:

$$\mathbf{M}\surd_s P(x_i) \text{ iff } s_i\in I(P)$$
$$\mathbf{M}\surd_s Q(a_i,x_j) \text{ iff } <I(a_i), s_j>\in I(Q)$$

When A does not contain occurrences of free variables, i.e. it is a sentence of L, then it is satisfied in **M** by one sequence s if and only if it is satisfied by all sequences. Hence, we can define **M**'s being a model of A, i.e. A is true in **M**, thus:

$$\mathbf{M}\surd A \text{ iff } \mathbf{M}\surd_s A \text{ for every } s$$

On the right hand side of each clausal biconditional you have a condition *with exactly the same logical form* as the sentence on the left, employing the same atomic formula structure for atomic sentences and the same connectives and quantifiers for the compound sentences; the corresponding facts are built up recursively matching the functional composition portrayed in the sentence-structure. Here we have a model of facts that preserves their intensional structure, thus immunizing it from the slingshot objection, and a perspicuous account of truth as correspondence with fact.[10]

(**C**) Sentence s is true in S iff there is a proposition p such that s designates p in S and p

(**C**) clearly resembles Tarski's T-schema but does have the advantage of making the semantic connection between sentences and their truth conditions explicit by the relation of designation. Niiniluoto says: "*[I]n this respect…the treatment of semantics by Carnap in the late 1930s and early 1940s was more satisfactory than Tarski's (1944) own explanations*." (Niiniluoto 1999, p. 96). While this is certainly true, it should not be overlooked that Tarski was explicit in restricting his attention to *interpreted* languages, i.e. languages assumed to be interpreted in the domain of all objects (see Tarski 1936, pp. 166-167 and Tarski 1969, p. 68).

[8] See, for example, Chang & Keisler 1973.

[9] We follow Niiniluoto 2002.

[10] This is how Tarski's theory meets Wittgenstein's (early) view that true sentences correspond to facts by being pictures of them. Both, I submit, understood that the logical structure of sentences is in every case a functional composition of corresponding names of Boolean functions. Thus, $\wedge$ names the Boolean meet, $\vee$ names the Boolean join, $\neg$ names complementation $\sim$, and $\forall$ and $\exists$ name the infinitary Boolean meet and join over all instances. The truth-table rules just ensure that the truth-value in the Boolean algebra {0,1} is given by the composite Boolean function mirrored in the structure of the proposition. For example, consider the formula '$p\wedge\neg(q\vee r)$'. Using 1 to symbolize 'true' and 0 to symbolize 'false', suppose $v(p) = 1$, $v(q) = 0$ and $v(r) = 0$ under a valuation v of the generators p, q and r. '$p\wedge\neg(q\vee r)$' is

London School of Economics and Political Science,
Department of Philosophy, Logic and Scientific Method
Houghton Street, LONDON WC2A 2AE
a.s.virdi@lse.ac.uk

References

Carnap, R. 1942. *Introduction to Semantics*, Cambridge, Massachusetts: Harvard University Press.
Chang, C. & Keisler, H. 1973. *Model Theory*, Amsterdam: North-Holland.
Church, A. 1943. Review of Carnap's *Introduction to Semantics*. *Philosophical Review* 52: 298-304.
Church, A. 1956. *Introduction to Mathematical Logic: Volume 1*, Princeton: Princeton University Press.
Davidson, D. 1969. True to the facts. In his *Inquiries into Truth and Interpretation*, 37-54. Oxford: Clarendon Press.
Davidson, D. 1990. The structure and content of truth. *Journal of Philosophy* 87: 279-328.
Gödel, K. 1944, Russell's mathematical logic. In Benacerraf & Putnam (eds.) 1964. *Philosophy of Mathematics: Selected Readings*, 447-469. Cambridge: Cambridge University Press.
Katz, J. and Fodor, J. 1963. The structure of a semantic theory, *Language* 39: 170-210.
Neale, S. 1995. The philosophical significance of Gödel's slingshot. *Mind* 104: 761-825.
Neale, S. and Dever, J. 1997. Slingshots and boomerangs. *Mind* 106: 143-168.
Neale, S. 2001. *Facing Facts*, Oxford: Clarendon Press.
Niiniluoto, I. 1999. Tarskian truth as correspondence - replies to some objections. In Peregrin (ed.) *Truth and its Nature (if any)*, 91-104. Dordrecht: Kluwer.
Niiniluoto, I. 2002. *Critical Scientific Realism*. Oxford: Oxford University Press.
Rodriguez-Peyera, G. 2003. Review of S. Neale, *Facing Facts*. *Mind* 112: 780-786.
Russell, B.A.W. 1905. On denoting. In his *Logic and Knowledge*, 41-56. London: Allen and Unwin Ltd.
Russell, B.A.W. 1918. The philosophy of logical atomism. In his *Logic and* Knowledge, 177-281. London: Allen and Unwin Ltd.
Tarski, A. 1936. The concept of truth in formalized languages. In Corcoran (ed.) 1983. *Logic, Semantics, Metamathematics: Papers by Alfred Tarski from 1922-1938*, 152-278. Indianapolis: Hackett.
Tarski, A. 1969. Truth and proof. *Scientific American* 220: 63-77.

---

true under the valuation v, i.e. $v(p\wedge\sim(q\vee r)) = 1$, just in case the Boolean function $1\wedge\sim(0\vee 0)$ has the value 1 in the Boolean algebra $\{0,1\}$, which it has. More generally, $v(p\wedge\sim(q\vee r))$ is the value of the function $v(p)\wedge\sim(v(q)\vee v(r))$.