Notes on Two Versions of the Slingshot Argument

Branden Fitelson
10/22/03

In these notes, I will provide relatively careful (annotated) renditions of two versions of the slingshot argument: Davidson's and Gödel's. Then, I will discuss a couple of possible replies to these arguments. In the end, I will conclude that the argument is not as compelling as it first appears, because it makes dubious (at least, controversial) assumptions about the theory of definite descriptions.

Davidson's Slingshot. First, some notation. I will use $\ulcorner(\hat{x})\phi\urcorner$ as shorthand for the definite description $\ulcorner$the x such that $\phi\urcorner$. I will use the letter "a" to denote an arbitrary existent concrete particular (*e.g.*, Cher). And, I will use "f_1" and "f_2" to denote arbitrary facts (*e.g.*, "snow is white" and "grass is green"). Finally, I will use "=" for identity, and "⇔" for logical equivalence.

1. The fact that f_1 = the fact that f_1.

   Remarks on (1). Nothing to say here. (1) is obviously true.

2. True identity statements *involving facts* remain true under the substitution of coreferential singular terms and/or logically equivalent statements.

   Remarks on (2). Note, (2) does *not* say that true identity statements *in general* remain true under substitution of equivalents/coreferentials. Consider the following identity involving *propositions*: (*) the proposition that snow is white = the proposition that snow is white. Perhaps we cannot expect (*) to remain true (in general) if we substitute coreferential singular terms. For instance, it turns out that white is my favorite color, but do we want to say that (**) the proposition that snow is white = the proposition that snow has my favorite color? Perhaps not. But, if not, then I presume this is because we have *identity conditions* for *propositions* that provide a rationale for denying (**). Those conditions might say, roughly, that in order for propositions to be identical, they must have the same truth-value in *all possible worlds*. And, my favorite color *could be* something other than white, which implies that there are bound to be possible worlds in which (intuitively) "that snow is white" is true but "that snow has my favorite color" is false. It is unclear why the identity conditions for *facts* should make (2) false. However, as I will explain below, there are independent reasons to worry about both aspects of (2).

3. (a) "That f_1" ⇔ (b) "that $(\hat{x})(x = a) = (\hat{x})(x = a \text{ and } f_1)$"

   Remarks on (3). We can prove (3), as follows. (⇒) Assume "that f_1" is true. Then, $(\hat{x})(x = a \text{ and } f_1) = (\hat{x})(x = a) = a$. So, "that $(\hat{x})(x = a \text{ and } f_1) = (\hat{x})(x = a)$" is true. (⇐) Assume "that f_1" is false. Then, $(\hat{x})(x = a \text{ and } f_1)$ is *empty*, but $(\hat{x})(x = a) = a$. So, "that $(\hat{x})(x = a \text{ and } f_1) = (\hat{x})(x = a)$" is false. Note: we are assuming here and throughout both arguments that a definite description $\ulcorner(\hat{x})\phi\urcorner$ stands for (*refers to*) *the unique thing satisfying* ϕ. This is *not* assumed by all theories of definitie descriptions (*e.g.*, Russell's, which does not assume that $\ulcorner(\hat{x})\phi\urcorner$ stands for or refers to *any* thing). See below, and [2] for further discussion. □

4. (c) "That f_2" $\Leftrightarrow$ (d) "that $(\hat{x})(x = a) = (\hat{x})(x = a \text{ and } f_2)$"

   Remarks on (4). The proof of (4) is exactly the same as the proof of (3), just with "f_2" substituted uniformly for "f_1". $\square$

5. (e) "$(\hat{x})(x = a \text{ and } f_1)$" and (f) "$(\hat{x})(x = a \text{ and } f_2)$" are coreferential.

   Remarks on (5). We can prove (5), as follows. Since f_1 and f_2 are *facts*, $(\hat{x})(x = a \text{ and } f_1) = (\hat{x})(x = a) = a = (\hat{x})(x = a \text{ and } f_2)$. Again, this assumes a *referential* reading of the operator $\ulcorner(\hat{x})\phi\urcorner$. $\square$

6. The fact that f_1 = the fact that $(\hat{x})(x = a) = (\hat{x})(x = a \text{ and } f_1)$.

   Remarks on (6). We can prove (6), as follows. By (3), (b) and (a) are logically equivalent. So, by (2), we can substitute (b) for (a) in the right hand side of (1), which yields (6). Intuitively, one might say, the *same fact* f_1 is what makes *both* "that f_1" *and* "that $(\hat{x})(x = a) = (\hat{x})(x = a \text{ and } f_1)$" true. But, it is unclear how intuitive this really is, as I will explain, below in the context of Gödel's rendition, which makes heavy use of this sort of inferential maneuver. $\square$

7. The fact that f_1 = the fact that $(\hat{x})(x = a) = (\hat{x})(x = a \text{ and } f_2)$.

   Remarks on (7). We can prove (7), as follows. By (5), (f) and (e) are coreferential. So, by (2), we can substitute (f) for (e) in the right hand side of the right hand side of (6), which yields (7). At this point, one is tempted to say something like the following. Wait a minute! It is fact f_1 that makes "that f_1" true, but it is fact f_2 that makes "that $(\hat{x})(x = a) = (\hat{x})(x = a \text{ and } f_2)$" true. But, this, of course, does not suffice to establish that $f_1 \neq f_2$. After all, if $f_1 = f_2$, then it will *still be true that* f_1 makes "that f_1" true and f_2 makes "that $(\hat{x})(x = a) = (\hat{x})(x = a \text{ and } f_2)$" true. What we need are *identity conditions for facts* that *tell us* whether $f_1 = f_2$. See below for discussion. $\square$

8. The fact that f_1 = the fact that f_2. [Hence, *there is at most one fact.*]

   Remarks on (8). We can prove (8), as follows. By (4), (c) and (d) are logically equivalent. So, by (2), we can substitute (c) for (d) in the right hand side of (7), which yields (8). $\square$

Gödel's Slingshot. Gödel's version of the slingshot argument is a bit longer, but it requires weaker assumptions (no general assumptions about logical equivalence) to reach a similar conclusion. Gödel's conclusion is that *if true sentences stand for facts, then they all stand for the same fact.* I will use "F" and "G" for arbitrary predicates, and "a" and "b" for arbitrary names (of particulars). Here, I am following Neale's [2] (*the* paper on the slingshot) very closely.

(1′) If $\ulcorner\phi(a)\urcorner$ and $\ulcorner a = (\hat{x})(x = a \text{ and } \phi(x))\urcorner$ are both true sentences, then they stand for *the same fact* (*i.e.*, the same fact "makes them" both true).

   Remarks on (1′). This may *seem* plausible [it's similar to the move made in (6) of Davidson's argument]. Intuitively, it may seem that it is "the fact that $\phi(a)$" which makes both $\ulcorner\phi(a)\urcorner$ and $\ulcorner a = (\hat{x})(x = a \text{ and } \phi(x))\urcorner$ true. But, this is not so clear. If $\ulcorner(\hat{x})\phi\urcorner$ is given a *referential* reading, then $\ulcorner(\hat{x})(x = a \text{ and } \phi(x))\urcorner$ refers to a (if it refers to anything). But, then, if true, $\ulcorner a = (\hat{x})(x = a \text{ and } \phi(x))\urcorner$ is (intuitively) 'made true by' the fact that $a = a$, whereas, if true, "$\phi(a)$" may not be 'made true by' that same fact (if $\phi(a)$ *is just* $a = a$, then it would be, but that cannot be assumed here). And, if $\ulcorner(\hat{x})\phi\urcorner$ is *not* given a referential reading (as in Russell's theory), then many of the inferences made later in this argument (and in Davidson's rendition of the argument, above) are not valid, since they assume that $\ulcorner(\hat{x})\phi\urcorner$ *refers* (to some particular). Thus, it seems there is a *dilemma* lurking here for the proponent of the slingshot – a dilemma arising from two different ways of thinking about definite descriptions $\ulcorner(\hat{x})\phi\urcorner$. $\square$

(2′) Assumption: "Fa" is a true sentence, which stands for the fact f_1.

(3′) Assumption: "$a \neq b$" is a true sentence, which stands for the fact f_2.

(4′) Assumption: "Gb" is a true sentence, which stands for the fact f_3.

(5′) "$a = (\hat{x})(x = a \text{ and } Fx)$" is a true sentence, which stands for fact f_1.

Remarks on (5′). It is clear that if "Fa" is a true sentence, then "$a = (\hat{x})(x = a \text{ and } Fx)$" is a true sentence. So, by (1′) and (2′), since "Fa" stands for f_1, so does "$a = (\hat{x})(x = a \text{ and } Fx)$". $\square$

(6′) "$a = (\hat{x})(x = a \text{ and } x \neq b)$" is a true sentence, which stands for fact f_2.

Remarks on (6′). It is clear that if "$a \neq b$" is a true sentence, then "$a = (\hat{x})(x = a \text{ and } x \neq b)$" is a true sentence. So, by (1′) and (3′), since "$a \neq b$" stands for f_2, so does "$a = (\hat{x})(x = a \text{ and } x \neq b)$". $\square$

(7′) "$a = (\hat{x})(x = a \text{ and } x \neq b)$" and "$a = (\hat{x})(x = a \text{ and } Fx)$" stand for the same fact. In other words, $f_1 = f_2$.

Remarks on (7′). This step assumes that (*i*) $\ulcorner(\hat{x})\phi\urcorner$ *refers* to the unique x satisfying ϕ, which allows us to infer that $(\hat{x})(x = a \text{ and } x \neq b) = (\hat{x})(x = a \text{ and } Fx) = a$; and (*ii*) that (quoting Gödel) "the signification [referent] of a composite expression, containing constituents which themselves have a signification [referent], depends only on the signification [referents] of these constituents (not on the manner in which this signification [referent] is expressed)." We have been assuming (*i*) all along, but Russell's theory of definite descriptions does not assume (*i*). See below and [2] for discussion – this is a promising response. (*ii*) seems to be in tension with (1′), since (1′) seems more plausible on a *non*-referential reading of $\ulcorner(\hat{x})\phi\urcorner$ (as I explained above), but (*ii*) seems silly if $\ulcorner(\hat{x})\phi\urcorner$ is non-referential. If $\ulcorner(\hat{x})\phi\urcorner$ is non-referential, then "the siginfication of $\ulcorner(\hat{x})\phi\urcorner$" is *empty.* $\square$

(8′) "$b = (\hat{x})(x = b \text{ and } Gx)$" is a true sentence, which stands for fact f_3.

Remarks on (8′). It is clear that if "Gb" is a true sentence, then "$b = (\hat{x})(x = b \text{ and } Gx)$" is a true sentence. So, by (1′) and (4′), since "Gb" stands for f_3, so does "$b = (\hat{x})(x = b \text{ and } Gx)$". $\square$

(9′) "$b = (\hat{x})(x = b \text{ and } a \neq x)$" is a true sentence, which stands for fact f_2.

Remarks on (9′). It is clear that if "$a \neq b$" is a true sentence, then "$b = (\hat{x})(x = b \text{ and } a \neq x)$" is a true sentence. So, by (1′) and (3′), since "$a \neq b$" stands for f_2, so does "$b = (\hat{x})(x = b \text{ and } a \neq x)$". $\square$

(10′) "$b = (\hat{x})(x = b \text{ and } a \neq x)$" and "$b = (\hat{x})(x = b \text{ and } Gx)$" stand for the same fact. In other words, $f_2 = f_3$.

> Remarks on (10′). The argument for (10′) is the same as the argument for (7′), which rests the assumption that definite descriptions refer, and on the assumption the referents of complex *facts* are determined solely by the referents of their constituents. As I explained above, these assumptions seem to be at odds with (1′). □

(11′) Since, $f_1 = f_2$ and $f_2 = f_3$, we have $f_1 = f_3$. So, "Fa" and "Gb" stand for the same fact. *Mutatis mutandis* where "$a = b$" (rather than "$a \neq b$") is true. Therefore, *all true sentences stand for the same fact.* □

Closing Remarks. There are only two places where this argument can really be challenged. First, its assumption that a definite description $\ulcorner(\hat{x})\phi\urcorner$ *refers to the unique thing x satisfying ϕ*. As Gödel stresses, this is *not* assumed in Russell's theory of descriptions. Moreover, on such a referential reading of $\ulcorner(\hat{x})\phi\urcorner$, assumptions like Gödel's (1′) and (the logical equivalence part of) Davidson's (2) are not very plausible [see also [2] and [1] for discussion]. Second, its assumption that (as Gödel puts it, my brackets) (†) "the signification of a composite expression [within the scope of a "*the fact that* ..."], containing constituents which themselves have a signification, depends only on the signification of these constituents (not on the manner in which this signification is expressed)." On a referential reading of definite descriptions, (†) seems plausible. But, on a non-referential reading, it is dubious, at best. Even on a referential reading of $\ulcorner(\hat{x})\phi\urcorner$, (†) seems problematic when applied to *propositions* (or *beliefs, etc.*), which are identified by their *profile of truth-values across all possible worlds*. Facts, on the other hand, don't seem to have such strong modal identity conditions. Intuitively, facts only *exist* in the actual world (and so they seem to have no "profile of truth-values across possible worlds"). As such, it is unclear why (†) should be false — as applied to *fact*-identities. It would be nice to have (independent) *identity conditions for facts* which provide a *rationale* for counterexamples to (†). This, as I see it, is the ultimate *metaphysical* challenge provided by the slingshot argument for the fact-theorist (if they believe there is more than one fact!). Russell's theory of facts (in which facts are structured ordered tuples of particulars and properties), together with his non-referential theory of descriptions, would certainly do the trick. That seems like one perfectly good way to avoid the slingshot. There are other ways to avoid the slingshot (see [1] for an alternative theory of facts and descriptions that avoids the slingshot). One thing seems clear: one cannot have *both* a *referential* theory of definite descriptions, *and* a purely *extensional* theory of facts (*i.e.*, a theory of facts on which facts do not change their identity across substitutions of coreferential terms). This, I think, is the general philosophical lesson of the slingshot.

References

[1] J. Barwise and J. Perry. Semantic innocence and uncompromising situations. In *The philosophy of language*, pages 369–381. Oxford Univ. Press, New York, 1996.

[2] S. Neale. The philosophical significance of Gödel's slingshot. *Mind*, 104:761–825, 1995.