The *solo numero* paradox

(forthcoming in the *American Philosophical Quarterly*)

István Aranyosi

(Bilkent University)

Leibniz notoriously insisted that no two individuals differ *solo numero*, that is, by being primitively distinct, without differing in some property. The details of Leibniz's own way of understanding and defending the principle –known as the principle of identity of indiscernibles (henceforth 'the Principle')—is a matter of much debate. However, in contemporary metaphysics an equally notorious and discussed issue relates to a case put forward by Max Black (1952) as a counter-example to any necessary and non-trivial version of the principle. Black asks us to imagine, via one of the fictional characters of his dialogue, a world consisting solely of two completely resembling spheres, in a relational space. The supporter of the principle is then forced to admit that although there are *ex hypothesi* two objects in that universe, there is no property (except trivial ones), not even relational ones, to distinguish them, and hence the necessary version of the principle is falsified.

In this essay I will argue that Black's possible world, together with the dialectic between the potential friends and foes of the Principle as expounded by Black himself and other authors, leads to paradox, which I will call 'the *solo numero* paradox'. That is not to say, however, that Black's world is itself not possible *per se*, but that, apparently, describing that world can never coherently settle the debate. I will offer a solution to the paradox, based on a new version of the principle, which, I will argue, is the weakest non-trivial version offered so far, and should be acceptable by both sides of the debate as close enough to the standard one. Black's world will be shown to verify this new principle.

1. Triviality and impurity

The Principle, stated in the most general form, that Black takes as his target can be formally expressed as follows:

$$(\mathrm{PII}_1)\ \Box[\forall x \forall y \forall \mathrm{P}\ (\mathrm{P}x \equiv \mathrm{P}y) \supset x=y]$$

That is: necessarily, for any individuals *x* and *y*, and any property P, if *x* has P if, and only if *y* has P, then x and y are identical. Or, more informally, if individuals *x* and *y* have exactly the same properties, then they are identical. A counterexample to (PII_1) is then of the following general form:

$$(\mathrm{C\text{-}PII}_1)\ \Diamond[\exists x \exists y \forall \mathrm{P}\ (\mathrm{P}x \equiv \mathrm{P}y)\ \&\ x \neq y]$$

Black's two-sphere world is intended to verify (C-PII_1), so to defend the principle is to deny either the indiscernibility (the first conjunct in the above formula), or the distinctness of *x* and *y*. Denying the second has not been without proponents (Ian Hacking 1975), but it is part of the present hypothesis that there really are *two* spheres in that world. Hence, the supporter of the principle is challenged to come up with a distinguishing property, i.e. a property such that it is not true that both *x* and *y* has it.

The first option considered by Black, and discussed extensively by Douglas Odegard (1964), is a so-called 'identity-property'[1], like *being identical to a*, where '*a*' is a name referring to one of the spheres. It is agreed by virtually everyone in the debate that such properties, to the extent that they can be considered properties to begin with, trivialize the principle. Of course, *being identical to a* will be a property that is not instantiated by *b*, under the assumption that there are two spheres and that '*b*' is a name for the sphere that is distinct from *a*. The problem is, as pointed out by several authors (Black 1952, Odegard 1964, Robert Adams 1979, Bernard Katz 1983, Gonzalo Rodriguez-Pereyra 2006), that allowing quantification over such properties in (PII_1) makes it trivial: to say that only *a* has the property of *being identical to a* is no different *in meaning* from saying that *a* and *b* are distinct, which was supposed to follow from

[1] Bernard Katz (1983) proposes the notion of *basic identity properties* (BIPs) in an attempt to explicate properties that trivialize PII_1: F is a BIP iff (1) it is possible that $\exists x(Fx)$ and (2) it is necessary that $\forall x \forall y(Fx \,\&\, Fy \ldots x = y)$. Predicates expressing BIPs are called 'BIP-predicates'. According to Katz, then, a predicate 'P' expresses a trivializing property iff 'P' contains a BIP-predicate essentially or 'P' may be defined in terms of some predicate that does. The notion of an identity property used in this paper is to be understood intuitively, as the property expressed by a predicate that essentially involves (a) the relation of identity and (b) at least one particular, definite object as its relatum. The two types of examples are: *being identical to a* and *being distinct from a*.

some claims regarding the properties instantiated by the two spheres, not to be assumed at the outset. So an argument for the principle based on considering such properties is both question-begging and proves nothing but a trivial version of it. Consequently, the quantifier binding the predicate variable in (PII_1) has to be restricted to some subset of properties, namely those that do not trivialize the principle.

A more inclusive category of properties that Black considers is that of impure properties. These are properties that essentially involve the existence of particular objects. Some of them are extrinsic relational, e.g. *being married to John Hawthorne*, *living close to the Eiffel Tower*. Others are intrinsic relational, e.g. *having Jay Leno's lower jaw as a proper part*. One question is whether such properties trivialize the principle. Some argue that they do (A. J. Ayer 1959: 26-35, Peter Forrest 2006), some that they don't (Rodriguez-Pereyra 2006). We should first notice that impure properties, *in general*, do not involve the danger of trivializing the principle. Consider an asymmetric world, where there are two intrinsically indiscernible spheres with a distance of 5 meters between them, and a cube, called '*c*', which is at 2 meters distance from one and 3 meters from the other. The property of *being 2 meters from c* is an impure property had by one of the spheres. Yet, the fact that one can ascribe this property to one variable without ascribing it to another variable shows that one can safely move to the existential instantiation of these variables by names, without any trivial assumption to the effect that there are two non-coreferring names[2].

[2] The names will be synonymous with the expressions 'the one that is two meters from *c*' and 'the one that is 3 miles from *c*'.

However, in the particular case of Black's universe any impure property has no chance but to involve either one or the other of the two spheres, hence, the problem in *that* universe seems to be that existential instantiation can never be legitimately appealed to. Existential instantiation is a valid inference rule only when the names that are used as instantiating the quantified formula are introduced precisely for the purpose of standing for the particular thing, whatever it may be, that makes the premise in quantified form true. However, for Black's universe such a strategy is not available. In that universe the following formula is presented as true, and, consequently, as a challenge to the supporter of the principle:

(1) $\exists x \exists y \forall \mathrm{P}\ (\mathrm{P}x \equiv \mathrm{P}y)\ \&\ x \neq y$

Since the second conjunct is unavailable for existential instantiation for the supporter of the principle, because it is supposed to be explained in terms of some facts about the first conjunct according to the principle, she has to somehow extract a property from the conceived situation which would distinguish two objects, *a* and *b*, e.g. by being ascribable to *a* and not to *b*. But the problem is that the very meaning of the quantified formula $\exists x \exists y \forall \mathrm{P}\ (\mathrm{P}x \equiv \mathrm{P}y)$ is such that the supporter of the principle can't even begin using two names, '*a*' and '*b*', in a legitimate way; given the first conjunct of premise (1), the only legitimate way to apply existential instantiation to it is:

$(2)\ \ \forall P\ (Pa \equiv Pa)$[3]

and not:

$(3)\ \forall P\ (Pa \equiv Pb)$[4]

Black himself (1952: 156-7) appeals to such considerations, but it is even better expressed by Odegard (1964: 205):

> "But, using 'A' and 'B' to refer to the *given* particulars is in effect an attempt to use different names to refer to each of them – a move which is *ex hypothesi* impossible because it contradicts the indiscernibility condition of the hypothetical case. For, the successful use of different names in this case presupposes the possibility of qualitatively distinguishing the given particulars, i.e., the possibility of saying truly 'A is *the* particular which …' and 'B is *the* particular which …'. And, *ex hypothesi*, there are no possible grounds for so distinguishing them. The exponent of the Principle must here prove the discernibility of the given particulars, not assume it; and the use of different names to refer to each of them simply assumes it. (…) If, on the other hand,

[3] To be sure, (2) is logically incompatible with the existential instantiation of $x \neq y$, but that is cold comfort for the supporter of the principle, for two reasons: (a) because she is asked to *derive* the instantiation of $x \neq y$, not to assume it, and (b) if she can't do that, the incompatibility would rather force her to deny $x \neq y$, which is contrary to one of the suppositions of Black's postulated universe.

[4] Of course, (3) itself does not discern *a* and *b*, but it can potentially serve as a premise for purposes of a *reductio* having as conclusion that (3) can't be true, i.e. that *a* and *b* are discernible after all.

> ‘A‘ and ‘B’ are not being used as different names in this context, then they must have something of the force of variables”. (*emphases as in the original*)

So the supporter of the principle is left with the choice of either using the names illegitimately, and hence beg the question, or to use them as disguised variables. The second option is bad news for her because now she can’t appeal to impure properties, e.g. *being at a nonzero distance from b,* which would distinguish the two spheres (since *a* is not at a nonzero distance from *a*), but can use only the pure versions entailed by the former, e.g. *being at a nonzero distance from something*. The impure versions can’t distinguish the spheres because it is true of both that: ’… is at a nonzero distance from something and at no nonzero distance from itself’.

2. Paradox

I agree that this is all bad news for the supporter of the Principle, who appears ensnared by her own theoretical commitments. However, if one focuses one’s attention on the resources the *opponent* of the Principle might deploy in accounting for Black’s universe, paradox ensues. Informally, the paradox is the following. The opponent of the Principle, as opposed to its supporter, is free and entitled to assume that the two objects are primitively distinct, that is, differ *solo numero*. So ‘*a* is distinct from *b*’ seems to be an unproblematic assertion on his part, as he can, and must by postulation, assume primitive distinctness. But that assertion being primitively true just means (i.e. biconditionally implies) that the above mentioned impure properties, e.g. *being at 5 meters distance from* *b*, are instantiated and will distinguish the two spheres. To be more precise, consider the relation, D, '… is at a nonzero distance from …'. Even though this relation holds symmetrically between *a* and *b*, still the impure versions of the unary relational properties that follow from D will be distinct properties, i.e., *being at a nonzero distance from a* and *being at a nonzero distance from b*. This is so because for the opponent of the Principle (as opposed to the supporter of it) the names '*a*' and '*b*' are not disguised variables. So even if both individuals have the property of *not being at a nonzero distance from itself*, still the relational properties involving nonzero distance from the other given object(i.e., *being at a nonzero distance from a* and *being at a nonzero distance from b*) are not instantiated by both. Hence they are discernible. But if they are discernible, then they are not *solo numero* different.

Now, for the formal version we will use some rules and notation of lambda calculus. A *lambda abstract* will have as its general form '(λv)(E)', where 'v' stands for a variable, 'λ' stands for the lambda-operator, and 'E' is called 'the body of the abstraction'. There are three rules that we will use, α-conversion and β-conversion, and λ-abstraction. α-conversion allows changing the names of λ-bound variables, i.e. replacing each of them with an arbitrary name. β-conversion is the operation of applying the function (i.e. the abstraction) to an argument; the result of such an operation is called an 'application'. The result of λ-abstraction is a lambda abstract, and can be thought of as extracting from a formula a singular term standing for a property. To give some intuitive examples, $(\lambda x)(\mathrm{R}xz)$ is α-equivalent to $(\lambda y)(\mathrm{R}yz)$, and $(\lambda n)(n+3)a$ β-equivalent to $a+3$.

From a formula like, e.g., $\exists x \mathrm{R}xy$ we can obtain by λ-abstraction $(\lambda x)(\mathrm{R}xy)$. Our first premise is that *a* and *b* are different *solo numero*:

(4) $\forall \mathrm{P}\ (\mathrm{P}a \equiv \mathrm{P}b)\ \&\ a \neq b$

The second premise is:

(5) $[\forall \mathrm{P}\ (\mathrm{P}a \equiv \mathrm{P}b)\ \&\ a \neq b] \supset \neg[(\lambda x)(\mathrm{D}bx)(a) \equiv_{\alpha} (\lambda x)(\mathrm{D}ax)(b)]$

What proposition (5) asserts is that *a* and *b* being different *solo numero* implies that the application to *a* of the property *being at a nonzero distance from b* is not α-equivalent to the application to *b* of the property *being at a nonzero distance from a*. The argument for (5) would be the following:

Suppose $a \neq b$. Consider the two objects in Black's world being at a certain nonzero distance from each other. Then $(\lambda x)(\mathrm{D}bx)(a)$ and $(\lambda x)(\mathrm{D}ax)(b)$. Also: $\neg(\lambda x)(\mathrm{D}bx)b$, and $\neg(\lambda x)(\mathrm{D}ax)a$. The latter two formulae can be expressed (via β-conversion, then λ-abstraction and α-conversion) as $(\lambda z)(\neg \mathrm{D}zz)a$ and $(\lambda z)(\neg \mathrm{D}zz)b$. That is, the property of not being at a nonzero distance from oneself is shared by both individuals. However, since the given objects are assumed as primitively distinct, the former two properties, involving the given particulars essentially, have to be distinct as well, namely, $(\lambda x)(\mathrm{D}bx)$ and $(\lambda x)(\mathrm{D}ax)$. This means that $(\lambda x)(\mathrm{D}bx)(a)$ and $(\lambda x)(\mathrm{D}ax)(b)$ are not α-equivalent.

From (5) and the second conjunct of (4) it follows that there is at least a property that is not shared by *a* and *b*, hence they are not *solo numero* different:

(6) $\exists$P $\neg$(P*a* $\equiv$ P*b*) (4, 5)

(7) $\neg\forall$P (P*a* $\equiv$ P*b*) & *a* $\neq$ *b* (6)

(8) $\perp$ (5, 8)

So if *a* and *b* are different *solo numero*, then they are not different *solo numero*.

The opponent of the Principle might try to find a property abstract that would apply to both *a* and *b*, such that it would somehow contain the two property abstracts present in the consequent of (5), but would not be reducible to them. If there were, e.g., an irreducibly disjunctive property having as disjuncts D*bx* and D*ax*, and shared by *a* and *b*, then the problematic impure properties (λx)(D*bx*) and (λx)(D*ax*) would not be applicable outside that disjunction to the given objects. The problem is that a property abstract like (λx)(D*bx* $\vee$ D*ax*), although applicable to both objects, it is not *irreducibly* disjunctive, given that it is known that neither of the objects instantiates D*xx*. In other words, (λx)(D*bx* $\vee$ D*ax*)(*a*) is reducible to (λx)(D*bx*)(*a*); *mutatis mutandis* for *b*. Hence, the implication in (5) holds just as before.

Another strategy could be to find a so-called arbitrary object *z* (also called ‘indefinite object’ in the literature) such that, e.g., *a* has the property of being the *z* such that *z* has the property of being either at a nonzero distance from *a* or from *b*, that is: (λz)[(λu)(D*bu* $\vee$ D*au*)(*z*)](*a*). And, of course, *mutatis mutandis* for *b*. Now, since *z* is an arbitrary object (e.g. whatever is denoted by ‘a sphere’, or ‘one of the spheres’ in the two-sphere universe) the body of the λ-abstract binding z is irreducibly disjunctive. In other words, $(\lambda u)(\mathrm{D}bu \vee \mathrm{D}au)(z)$ is not reducible to either $(\lambda u)(\mathrm{D}bu)(z)$ or $(\lambda u)(\mathrm{D}au)(z)$. However, as pointed out by Kit Fine (1983: 62), β-conversion fails in the case of disjunctions as applied to arbitrary objects. More precisely, the rule of disjunction is not applicable to arbitrary objects in any direct way. Consider an arbitrary natural number. The range of the arbitrary object is the set of natural numbers. Some are odd, some are even. So we can assert that the arbitrary object is odd or even. But an arbitrary natural number is not even, and it is not odd. So the disjunction rule (i.e. that a disjunction is true iff some of its disjuncts is true) fails.[5] Hence, as applied to our case: $\neg[(\lambda z)[(\lambda u)(\mathrm{D}bu \vee \mathrm{D}au)(z)](a) \equiv_\beta (\lambda u)(\mathrm{D}bu \vee \mathrm{D}au)(a)]$.

3. Instantial terms, arbitrary objects, and context dependent quantifiers

The above discussion seems to bring about a trilemma. The supporter of the Principle has to choose between taking ‘*a*’ and ‘*b*’ as having the force of variables or that of proper names designating particular objects. As made clear above, the first choice doesn’t account for the distinctness of the two spheres in Black’s world, whereas the second was not available to her to begin with, because it presupposed that the objects had already

[5] The only sense in which we have a disjunction here is indirect, namely, that we can apply the disjunction rule to *each* of the individuals that are in the range of the arbitrary object, *viz.* $(\lambda x)(\zeta x \vee \psi x)(x) \equiv_\beta \forall i\ (\zeta i \vee \psi i)$, where $(1, \dots\ i, \dots, n)$ is the range of x.

been distinguished. Thirdly, the opponent of the Principle can appeal to ‘*a*’ and ‘*b*’ as names of primitively distinct objects, but then he is forced into paradox, because now impure properties of the type *being at a distance from a* will be applicable to each object, and will distinguish them, hence they won’t be primitively distinct.

The solution is based on denying that the choice between variables and proper names as interpretations for the terms ‘*a*’ and ‘*b*’ is a genuine dichotomy. There is a third category that these terms can be taken to be, namely, what Jeffrey King (1991) calls ‘instantial terms’. These are terms that in mathematics are used to designate an arbitrary mathematical object (as in ‘Take an arbitrary prime number *n*’; *n* is the instantial term), and in the logic of natural deduction are free variables introduced by the rule of existential instantiation (as in ‘There are some red things ($\exists x \mathrm{R}x$). Take one of them ($\mathrm{R}y$). …’; *y* is the instantial term) and eliminated by the rule of universal generalization (as in ‘Take an arbitrary person ($\mathrm{P}x$). It is tall ($\mathrm{T}x$). Hence, all persons are tall ($\forall y\ \mathrm{P}y \supset \mathrm{T}y$)’; *x* is the instantial term).

However, it is not the syntax that is interesting about such terms, but their semantics and the truth conditions associated with their use. Fine (1983, 1985a, 1985b) offers a theory of arbitrary objects as the designata of such instantial terms, offering, among other things, individuation conditions for them, whereas King (1991) argues that they are disguised context dependent quantifiers (CDQ), where the context encodes all the relevant information for the properties of the instantial term within a derivational structure. We need not enter all the details of these accounts, as for all our purposes they will both give the very same verdicts, but focus on two important aspects. First, we should point out that instantial terms have a generality that makes them unlike terms that refer to a definite object, like, e.g., proper names do. But unlike variables, they are never supposed to be bound by quantifiers. And, second, and this is most emphasized by King, there are dependence relation between the semantic value of these terms and the derivational context in which they figure.

The idea of context dependence is that truth of a sentence with an occurrence of an instantial term is always evaluated in a derivational context, i.e. in a formal argument, or in an argument using ordinary English. Let us use an example offered by King (1991: 246).

(Fl)

premise (1) Every professor has a bad student.

premise (2) Every bad student hates each of his/her professors.

(3) Consider an arbitrary professor.

(4) By (1), <u>the professor</u> has a bad student.

(5) Consider <u>the professor's bad student</u>.

(6) By (2), <u>the student</u> hates <u>the professor</u>.

(7) So every professor is hated by some student.

There are various dependence relations here. According to King’s CDQ-account, both types the underlined terms, ‘the professor’ and ‘the student’, are context dependent quantifiers. Their force (existential or universal) depends on what propositions they were derived from and what other propositions have been assumed. For instance, the force of ‘the professor’ in (4) is universal, as it is derived from (1) where the variable was bound by a universal quantifier. The force of ‘the professor's bad student’ in (5) is existential, as the clause ‘Consider …’ in (5) is dependent on an existential quantifier in (4), which in turn depends on the existential quantifier applied to ‘student’ in (1). Intuitively, having already considered an arbitrary professor from all the professors, we have thereby considered that the professor has some bad student. So considering that bad student is going to preserve the quantificational force of ‘some student’.

Similarly, their scopes relative to each other are derivationally context dependent. For instance, the scope of ‘the student’ in (6) is wider relative to that of ‘the professor’. Finally, the domains over which they range are also in a dependence relation, as the clause ‘Consider ...’ in (5) determines the domain associated with the subsequent occurrences of ‘the student’.

Fine (1983) provides a similar story, based on dependent and independent arbitrary objects, and offers individuation criteria. Independent arbitrary objects are identical iff they have the same range. Dependent ones are identical iff they have the same range, and depend on the same arbitrary objects in the same way. A useful analogy here is with independent versus mutually exclusive events in statistics. For instance, the lottery involves a number of random draws of natural numbers from a finite set, such that after the first number is selected all subsequent draws depend on the previous choice in the sense that the previous numbers are not in the set any more. If after each draw the number were put back in the pool, the draws would be independent. Things are similar with instantial terms if they stand for arbitrary objects. There are derivational contexts in which two or more instantial terms are independent, and there are such contexts in which the reference of some depends on what choices of arbitrary objects have already occurred within the derivation. The above derivational structure, (F1), can be analyzed in terms of dependent arbitrary objects just as in terms of CDQs. The difference between King's and Fine's approach is that the former stresses derivational context dependence as an essential feature of instantial terms, whereas the latter mentions this as more like a 'side-effect' of his theory of arbitrary objects.

4. Back to the Principle

The above discussion is relevant to the debate about the Principle in the following way. Taking '*a*' and '*b*' as instantial terms when inquiring about whether the Principle has counter-examples or not has a number of positive consequences, for both the supporter and the opponent of it.

Let us first consider the supporter. The evaluation of the Principle's truth value will involve instantial terms, and hence will essentially involve a context – a derivational structure or any relevant discourse fragment within which the evaluation is undertaken. The context will determine for each occurrence of such terms their reference, considering them as referring to arbitrary objects, or their associated quantifier, under the CDQ approach. The novel feature of the Principle is now that we don't need to include impure properties, which essentially involve definite objects, in the domain of quantification associated with the predicate variable in (PII_1), as they are controversial as far as their potential threat to trivialize the Principle is concerned. Instead, the domain will include only pure properties, but some of these will be what we will call 'instantial properties', i.e., the derivational context dependent properties that the variables get by being instantiated by the instantial terms, given a derivation. Given this, the new principle will explicitly mention derivational context:

$$(\mathrm{PII}_2)\ \Box[\forall x\forall y\forall \mathrm{P}\forall C\ (\mathrm{P}x \text{ in } C \equiv \mathrm{P}y \text{ in } C) \supset x=y]$$

The new principle states that necessarily, if *x* and *y* are indiscernible in all derivational contexts, then they are identical, or that no two things can possibly be indiscernible in all derivational contexts. Indiscernibility in a derivational context will involve, of course, sameness of standard instrinsic and extrinsic (pure) properties, but it will also involve, as stated above, sameness of instantial properties – sameness of force, range, and relative scope for the relevant instantial terms introduced or eliminated during the derivational processes involving quantified formulae. The novelty of the approach is to introduce, besides the standard (pure) properties that object might have, the (pure) properties these objects have *qua* objects of reasoning.

The new principle seems true: if some objects are indistinguishable both in their pure properties and in our reasoning about them, then they are one and the same object. Black's universe, or any similarly symmetric universe (e.g. with three spheres arranged as an equilateral triangle), are not counter-examples to (PII_2) since there are derivational contexts in which the instantial terms have different relational properties relative to the context. For instance, in the following derivational context associated with the two-sphere world:

(C_1)

(1) All spheres are at a distance from a sphere ($\forall x \exists y$ S$x \supset$ Dxy & Sy). (*hypothesis*)

(2) Consider an arbitrary sphere (Sz).

(3) <u>It</u> is at a distance from a sphere ($\exists y$ Dzy & Sy).

(4) Consider <u>the sphere that the arbitrary sphere is at a distance from</u> (Su & Dzu).

(5) Suppose <u>it</u> is made of iron (Iu).

(6) …

Wherever the particular derivation may further proceed, we can already extract information to the effect that the arbitrary object that is the reference of ‘it’ in (3) is different from that which is the reference of ‘it’ in (5). Taking them as CDQs, the range of z is all the spheres, whereas the range of u is restricted by the ‘consider..’ clause in (4), so that it includes everything except the arbitrary sphere z. Such instantial properties, therefore, make us able to distinguish the two spheres. It is important to note that any relation or relational property will do as predicated by an arbitrary object when it comes to distinguishing the objects via instantial properties, except the relation of identity, or a relational property involving identity. In that case, the substitutivity of identity will make it the case that one can, and has to, use the same instantial term when considering the arbitrary object that stands in the identity relation with some previously selected arbitrary object. So, for instance, if in the above derivational context we replace *being at a distance from something* with *being identical to something*, then we get:

(C_2)

(1) All spheres are identical to a sphere ($\forall x \exists y\ Sx \supset x = y\ \&\ Sy$). (*hypothesis*)

(2) Consider an arbitrary sphere (Sz).

(3) It is identical to a sphere ($\exists y\ z = y\ \&\ Sy$).

(4) Consider the sphere that the arbitrary sphere is identical to (Sz).

(5) Suppose it is made of iron (Iz).

(6) …

Clause (4) must be taken as a duplicate of clause (2), as the predicate 'is identical to' makes the substitution 'z' for 'y' in (3) legitimate given the substitutivity of identity. On the other hand, the relation of *being distinct from* stated in the hypothesis would make the distinctness between the two arbitrary objects trivial, as it is already assumed in the hypothesis, as a clause like 'Consider the sphere that is distinct from the arbitrary sphere' would by itself, analytically, entail that a new instantial term has to be introduced, rather than merely as a function of the context containing previous such clauses. But excluding both the identity property and impure properties from such derivational context is a safe way to keep the Principle safe from trivialization.

So, taking the names for the two spheres as instantial terms makes it possible in a non-trivial way for the Principle not to be falsified by Black's world. Clauses of the type 'Consider an arbitrary sphere' are constituents of many derivational contexts that can distinguish the two spheres. Black himself, at one point, considers such a clause, via his character *A*, but swiftly –too swiftly, indeed– rejects it; the passage is this (Black 1952: 156):

> " ***A***. (…) Consider one of the spheres, *a*, …
>
> ***B***. How can I, since there is no way of telling them apart? *Which* one do you want me to consider?
>
> ***A***. This is very foolish. I mean either of the two spheres, leaving you to decide which one you wished to consider. If I were to say to you "Take any book off the shelf" it would be foolish on your part to reply "Which?".
>
> ***B***. It's a poor analogy. I know how to take a book off a shelf, but I don't know how to identify one of two spheres supposed to be alone in space and so symmetrically placed with respect to each other that neither has any quality or character that the other does not also have."

The error here in *B*'s reasoning is to assume that the name '*a*' in 'Consider one of the spheres, *a*, …' must be taken as a constant designating a definite object, instead of, quite naturally, being taken as an instantial term designating an indefinite object, or a CDQ. *B* seems to be committed to the view that to consider an arbitrary sphere implies to consider a definite sphere, i.e., in *B*'s words, to 'know how to identify one of two spheres', by which he means to select a sphere that has been *individuated* beforehand. There is a simple and intuitive argument why this is not correct. When in mathematical proofs we are asked to consider a prime, all we are supposed to know is the definition of a prime number, i.e. the property that defines the class of primes. If knowledge of all definite primes were required, then we would either never be able to consider a prime, or considering a prime would mean considering a number that we know to be prime and can distinguish it from all other primes. Both of these options is absurd. It obviously makes sense to consider a prime, and the range of primes is much larger than the range of numbers that are known to be primes and are known to be distinguished from other primes. Considering a prime is compatible with the considered prime never having been thought about, as a definite number, by anyone, let alone known to be a prime.

Further evidence that it is definite objects that character *B* assumes to be needed when being asked to consider a sphere in Black's world is apparent in the lest lines of the part of the dialogue on this topic, when we uses the analogy of a mathematician who thinks that the Axiom of Choice ensures one the ability to choose a definite object (Black 1952: 157):

> " ***A***. All I am asking you to do is to think of one of your spheres, no matter which, so that I may go on to say something about when you give me a chance.
>
> ***B***. You talk as if naming an object and then thinking about it were the easiest thing in the world. But it isn't so easy. Suppose I tell you to name any spider

> in my garden: if you can catch one first or describe one uniquely you can name it easily enough. But you can't pick one out, let alone "name" it, by just thinking. You remind me of the mathematicians who thought that talking about an Axiom of Choice would really allow them to choose a single member of a collection when they had no criterion of choice".

Again, it is clear from this passage that character *B* is thinking in terms of proper names designating definite objects when asked to consider an arbitrary object. But this is not what is required: all one needs is what has been called instantial terms, or temporary constants (Patrick Suppes [1957] 1999: 81), or dummy-names (E. J. Lemmon 1961: 253), which stand for either an arbitrary object, or, if that theory is not acceptable[6], for a CDQ or some similar item. What is important about these names is that they are used temporarily in truth-preserving derivational structures, for the very purpose of ensuring truth-preservation. Further, if what we have pointed out in this section is right, then they do also play a role in non-trivially discerning, by their feature of context-dependence, the relevant objects that they designate.[7]

The appeal to instantial terms and, consequently, the introduction of instantial properties as properties of objects in their role in reasoning might seem somewhat ad hoc.

---

[6] The history of repulsion vis-a-vis arbitrary objects is long, indeed, probably starting with Berkeley's attack on abstract ideas. Nothing hinges in this paper on whether one accepts arbitrary objects or not; all we require to be admitted is that proper names designating definite objects are not at issue in the Black type universe, when it comes to evaluating the Principle.

[7] Leaving aside that nothing like the Axiom of Choice needs to be assumed in the two sphere universe (since the axiom is only required when the choice set is infinite) in order to talk about an arbitrary sphere, still, nothing beyond the kind of general knowledge similar to what is involved in reasoning based on that axiom is required to be able to refer to any one of the the spheres.

It is true that since Max Black's challenge the discussion has been run in terms of how to translate an ontological fact, like that of there being two distinct and perfectly alike spheres in a symmetric universe, into a logical language where this fact would be expressed without appeal to the relations of identity and non-identity, but Leibniz principle is not obviously about how we reason about objects, but rather about what objects are. Hector-Neri Castañeda (1975), for instance, makes a distinction between the ontological problem of individuation, which according to him has to do with the internal constitution of individuals, and the epistemological problem of individuation, which inquires into the ways in which reasoners are able to single out individuals and distinguish them from other individuals. Castañeda's point is that these two aspects of the problem of individuation have been many times run together as if they were the same, and that they deserve separate discussion.

In response I would point out that the approach in this paper is neither that of confusing the two problems, nor that of considering them as completely separate issues. Rather, the idea is to combine the two into a coherent picture. Second, it is not claimed that one should exclusively focus on properties that objects have in their role in reasoning, but rather that such properties should be allowed in one's repertoire when it comes to analyzing the relationships among objects in a domain of discourse, especially when which objects has what name is the very issue to be addressed. The idea is well exemplified by the dynamic approach to semantics, for instance in Jeroen Groenendijk, Martin Stokhof and Frank Veltman (1996). Dynamic semantics has been put forward as an alternative to standard truth-conditional semantics in order to account for a notion of meaning, present in natural language discourse, that depends on the dynamics of discourse (dialogue, reasoning, story-telling, etc.), that is on how informational states at each moment of an ongoing discursive process are being updated and with what semantic consequences. The meaning of a sentence in this setting is not its truth-conditions but the change in informational states that an assertion of that sentence can potentially bring about. Variables are taken as anaphors, while quantifiers can bind variables beyond their syntactic scope. What is relevant from the point of view of the present paper is that in dynamic semantics there are two equally important types of information: information about the world and information about the discourse. Although information about the discourse --for instance, about what has been said before in a derivation-- appears in some sense as less 'real', it is no less important than genuine information about the world when it comes to correctly accounting for several relations among objects *in the world*:

> "Discourse information of this type looks more like a book-keeping device, than like real information. Yet, it is a kind of information which is essential for the interpretation of discourse, and since the latter is an important source of information about the world, discourse information indirectly also provides information about the world." (Groenendijk, Stokhof and Veltman 1996: 183)[8]

5. Back to the paradox

[8] Thanks to an anonymous referee for the objection and the suggestion regarding dynamic semantics.

How is the opponent of the original Principle, (PII_1), to make use of this theory? One thing to note is that she need not be an opponent of the new Principle, (PII_2), as the thought behind the Black type universe as a case of difference *solo numero* assumes that, since the distinctness fact involving the spheres (i.e. the one we get by instantiating $x \neq y$) involves them as definite objects, the indiscernibility fact involving them (i.e. the one we get by instantiating $\exists x \exists y \forall P\ Px \equiv Py$) must also involve them as definite objects, whereas (PII_2) involves arbitrary objects or CDQs in its antecedent (and definite objects as the ones we get by instantiating its consequent). More importantly, the paradox we described as marring the very attempt to formulate the *solo numero* doctrine can now be avoided, because (PII_2) is compatible with the compossibility of multiple *definite* objects and only one arbitrary object or one CDQ , having the definite objects in its range. Such a compossibility is equivalent to there being distinct, independent possible derivational contexts, C_3 and C_4, such that considering an arbitrary sphere in C_1 and one in C_2 picks out the same arbitrary object, but might pick out distinct definite objects.

To see that this is so, note that the arbitrary objects or CDQs considered in such distinct derivational contexts will function as independent from each other. Since for independent arbitrary objects or CDQs the identity criterion is the sameness of their range, if the two derivational contexts are duplicates, the arbitrary object picked out in each context will be the very same, because duplicate derivational contexts will have the same domain of quantification and the same range for their potential instantial terms. At the same time, of course, the range of this arbitrary object or CDQ can contain any number of definite object. To exemplify, consider the following two duplicate derivational contexts involving Black's universe:

C_3

1. There are two spheres made of iron.

   $(\exists x \exists y\ \mathrm{S}x\ \&\ \mathrm{S}y\ \&\ \mathrm{I}x\ \&\ \mathrm{I}y\ \&\ x \neq y)$

2. Consider one of them. $(\mathrm{S}z)$

3. It is made of iron. $(\mathrm{I}z)$

4. …

C_4

1. There are two spheres made of iron.

   $(\exists x \exists y\ \mathrm{S}x\ \&\ \mathrm{S}y\ \&\ \mathrm{I}x\ \&\ \mathrm{I}y\ \&\ x \neq y)$

2. Consider one of them. $(\mathrm{S}w)$

3. It is made of iron. $(\mathrm{I}w)$

4. …

Here, although they are bearing different names, i.e. 'z' and 'w', the arbitrary objects picked out by the 'Consider..' clause in the two derivational context are identical, because they are independent of each other. But, of course, considering an arbitrary sphere in C_3 and one in C_4 does not tell us whether, if asked to think in terms of definite spheres, we have picked out the very same definite sphere or distinct such spheres. Still, the unique arbitrary sphere that is considered in both contexts has a range of two definite spheres, *ex hypothesi*. It is in this sense that the opponent of the PII_1 was right in thinking that there can be objects that are different *solo numero*, namely, the case in which we do not keep the derivational context fixed as required by PII_2. This means that PII_2 is a principle that should equally be acceptable to both the opponent and the supporter of PII_1.

The reason the *solo numero* paradox ensued was proposition (5), restated here:

$$(5)\ [\forall \mathrm{P}\ (\mathrm{P}a \equiv \mathrm{P}b)\ \&\ a \neq b] \supset \neg[(\lambda x)(\mathrm{D}bx)(a) \equiv_{\alpha} (\lambda x)(\mathrm{D}ax)(b)]$$

This proposition does not mention any derivational context as required by PII_2 for asserting the truth or otherwise of formulae containing '*a*' and '*b*'. If we hold fixed such a context across the lines of comparison between *a* and *b* in terms of the property abstracts that are applied to them, then PII_2 is verified by the new proposition (5*):

$$(5^*)\ [\forall \mathrm{P} \forall C\ (\mathrm{P}a \text{ in } C \equiv \mathrm{P}b \text{ in } C)\ \&\ a \neq b] \supset \neg[(\lambda x)(\mathrm{D}bx)(a) \equiv_{\alpha} (\lambda x)(\mathrm{D}ax)(b)]$$

If, on the other hand, we allow for variation in context, then the *solo numero* doctrine finds its place in paradox-free manner:

$$(\mathrm{SN})\ \forall \mathrm{P} \exists C_1 \exists C_2\ (\mathrm{P}a \text{ in } C_1 \equiv \mathrm{P}b \text{ in } C_2)\ \&\ a \neq b$$

This proposition is paradox-free because it does not imply that the *b*-involving property abstracts as applied to *a* are not α-equivalent to the *a*-involving property abstracts as applied to *b*. So, (5**) is false:

$$(5^{**})\ \mathrm{SN} \supset \neg[(\lambda x)(\mathrm{D}bx)(a) \equiv_{\alpha} (\lambda x)(\mathrm{D}ax)(b)]$$

(5**) is false because it is not the case that property abstracts of the form $(\lambda x)(\mathrm{D}bx)$ and $(\lambda x)(\mathrm{D}bx)$ apply to *a* and *b*, respectively, taken as definite objects, across derivational contexts. They could only apply within the same derivational context, but in that case, of course, given our background theory underlying PII_2, the difference is not *solo numero*.

To conclude, there is a way to understand the assertion that objects can be different *solo numero* that doesn't lead to paradox – it is that two or more definite objects can have the very same pure properties, including instantial ones, given that they are reasoned about in distinct derivational contexts. On the other hand, there can be no two definite objects that have the same pure properties, including instantial ones, if these objects are reasoned about within the same derivational context.

6. A final note: difference *solo nomine*?

In a brief note, Nicholas Rescher (1955) comes close to the main idea behind the present theory which underlies PII_2, when he writes:

> "As a starting point we accept the statement that 'The principle of the identity of indiscernibles may be taken to mean that if two objects O_1 and O_2 are numerically different then they are qualitatively different, they differ in some mentionable *respect*'. The word 'mentionable' deserves special scrutiny; it contains the version of the principle which it is the object of this paper to examine, for it establishes the role that *discourse* plays in the principle. From this viewpoint the principle of the identity of indiscernibles is not ontological (dealing with things that are or might be), nor, *a fortiori*, physical (dealing with

> the natural phenomena of the world about us). Rather, the contention which the principle makes, in this interpretation, is *semantic*; (…)” (1955: 153)

Unfortunately, Rescher does not come to recognize the role context can play in discerning objects, and does not develop his idea beyond a general suggestion to the effect that the Principle holds when the language is rich enough to contain uniquely referring expressions for each distinct object in its domain of reference:

> “ In this interpretation principle of the identity of indiscernibles asserts that any two objects in the intended domain of reference of language (i.e. a language) which are *in fact* different can be distinguished *in that language* (…).” (1955: 153) (*emphases as in the original*)

There are various problems with Rescher’s proposal, which there isn’t enough space here to discuss, but one that is apparent might be put as follows. Rescher’s criterion for evaluating the truth-value of the Principle will depend in Black type cases on purely metalinguistic considerations. It won’t appeal to predicates like ‘is red’, or ‘is made of iron’, etc., because these are assumed as being shred by both objects, if shared by any. So the relevant predicates will be metalinguistic, of the kind: ‘is called ‘a’ in language *L*’, or even ‘is the denotation of ‘the sphere that is at a distance from the one that is called ‘a’ in language *L*’ in language *L*’.

The problem with this approach is that now the idea of difference *solo numero* is replaced with the possibility of difference *solo nomine*, i.e. objects differing solely in name and not in properties. To see this point consider two languages, *L* and *L**, such that the only difference between them is that the references of names '*a*' and '*b*' are switched between them. Now it is hard to make sense of the idea that each language by itself can discern the two objects without assumptions about the referential relation between language and world. The two languages are *per se* indiscernible; there is no more sense in asserting that the objects are distinguished 'in that language' – which one, we may ask? The only way to even begin to discern them is to check out the referential relations between their terms and the domain, but that is not allowed by Rescher's assumption that the Principle is 'not ontological', but only 'semantic' (actually, metalinguistic in the relevant cases, i.e. the cases of potential counter-examples).

To round up the argument, our proposal does not have the problem of *solo nomine* difference. The new principle, PII_2, is both semantic, in some sense, and ontological, in another. It is semantic in that it essentially involves referential dependence relations among temporary names, i.e. instantial terms, within a given derivational structure. But it is also ontological in that these context-determined roles for instantial terms are taken as bona fide properties, i.e. instantial properties, of the objects in question. One can even go further and analyze these instantial properties in a context in terms of relations to the brain states of a reasoner when effecting the derivation that constitutes the context. Further, as we have seen when discussing the way to solve the *solo numero* paradox, from the point of view of the properties of arbitrary objects or CDQs there are no distinct but indiscernible derivational contexts; when assuming two distinct and duplicate such contexts, i.e. C_3 and C_4 above, we were led to accept a unique assignment of arbitrary objects or CDQs across them, because identity of range in the case of independent arbitrary objects or CDQs entails their identity (although, of course, these arbitrary objects or CDQs might have multiple definite objects in their range). In sum, the question 'which one is which?' does not arise in the case of C_3 and C_4, as it does in that of Rescher's languages L_1 and L_2 simply because the references and structural properties of the relevant instantial terms do not vary across such duplicate contexts.

7. Conclusion

Consequently, a good case can be made for the present approach to the Principle, both in terms of its internal coherence, which has been the task to fulfill in this essay, and in terms of potential application to special problems related to the Principle. Such special problems include the recently debated issue of how to accommodate a bundle theory of objects with Black type universes, or whether facts described by quantum mechanics are in conflict with the Principle, thus falsifying it even in of its modally weak forms, that of it being true of the actual world. The present theory might be able to deal with these issues, and might have various advantages over the extant approaches, but this task has to be put aside for another occasion.[9]

References

Adams, R. M., 1979, 'Primitive Thisness and Primitive Identity', *Journal of Philosophy*, 76: 5-26.

Ayer, A. J., 1959, *Philosophical Essays*, London: Macmillan.

Castañeda, H.-N., 1975, 'Individuation and Non-identity: a New Look', *American Philosophical Quarterly*, 12: 131–140.

Fine, K., 1983, 'A Defence of Arbitrary Objects', *Proceedings of the Aristotelian Society Supplementary Volume*, 57: 55-77.

Fine, K., 1985a, 'Natural Deduction and Arbitrary Objects', *Journal of Philosophical Logic*, 14: 57-107.

Fine, K., 1985b, *Reasoning with Arbitrary Objects*, Aristotelian Society Series Volume 3, Basil Blackwell, Oxford.

[9] Conversations with Tufan Kıymaz, Ezgi Ulusoy Aranyosi, Sun Demirli, and Hilmi Demir prompted me to think more about problems of individuation of particulars; I would like to thank them for these. I would also like to thank an anonymous referee for helpful suggestions, as well as to the Scientific and Technological Research Council of Turkey (TÜBITAK) for continued support of my research. I dedicate this paper to the memory of József Aranyosi, my father.

Forrest, P., 2006, 'The Identity of Indiscernibles', *The Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/entries/identity-indiscernible/

Groenendijk, J., M. Stokhof, and F. Veltman, 1996, 'Coreference and Modality' In: S. Lappin (ed.), *Handbook of Contemporary Semantic Theory*, Oxford, Blackwell, 1996: 179-216.

Hacking, I., 1975, 'The Identity of Indiscernibles', *Journal of Philosophy*, 72: 249-256.

Katz, B., 1983, 'The Identity of Indiscernibles Revisited', *Philosophical Studies*, 44: 37–44.

King, J. C., 1991, 'Instantial Terms, Anaphora and Arbitrary Objects', *Philosophical Studies* 61: 239-265.

Lemmon E. J., 1961, 'Quantifier Rules and Natural Deduction', *Mind* 70: 235-238.

Odegard, D., 1964, 'Indiscernibles', *Philosophical Quarterly*, 14: 204-213.

Rescher, N., 1955, 'The Identity of Indiscernibles: A Reinterpretation', *Journal of Philosophy*, 52: 152-155.

Rodriguez-Pereyra, G., 2006, 'How Not to Trivialise the Identity of Indiscernibles'. In P. F. Strawson and A. Chakrabarti (eds.) *Concepts, Properties and Qualities*, Ashgate, 2006: 205-223.

Suppes, P., [1957]1999, *Introduction to Logic*, New York: Dover Publications.