

On a Misguided Argument for the Necessity of Identity

Ari Maunu

University of Turku

Department of Philosophy

FI-20014 TURKU

FINLAND

arimaunu@utu.fi

<http://users.utu.fi/arimaunu/>

Orcid 0000-0003-0259-5493

Abstract

There is a certain popular argument, deriving from Ruth Barcan and Saul Kripke, from the conjunction of the Principle of the Indiscernibility of Identicals (PII, for short) and the Principle of the Necessity of Self-Identity to the Thesis of the Necessity of Identity. My purpose is to show that this argument does not work, not at least in the form it is often presented. I also give a correct formulation of the argument and point out that PII is not even needed in the argument for the necessity of identity.

Keywords: necessity, identity, rigidity, substitutivity

I. INTRODUCTION

According to a popular argument for the necessity of identity, deriving from Barcan 1947, Marcus 1961,¹ and Kripke 1971, actual identity of an object a with an object b entails necessary identity of a with b , because on the assumption that $a=b$ the necessity of b 's being identical with a follows from the self-evident necessity of a 's being identical with a by the Principle of the Indiscernibility of Identicals (aka, "Leibniz's Law"), or "whatever is true of a thing is true of anything identical with that thing" (PIInI, for short).

After presenting in Section II E. J. Lowe's (2002) version of the argument for the necessity of identity (which I shall call "the putative argument"), I show in Section III that the reasoning represented by his formulation is mistaken (as an attempt to prove the necessity of identity). In Section IV I present the correct version of the argument in question. Finally, in Section V I point out that PIInI is not even needed in a conclusive argument for the necessity of identity.

II. THE PUTATIVE ARGUMENT FOR THE NECESSITY OF IDENTITY

Taking E. J. Lowe's (2002: 84–90)² presentation of the argument for the necessity of identity as an example, Lowe gives the following formulation of this argument (Lowe 2002: 85):

Suppose [...] that a is identical with b , where a and b are any objects whatever. Now, by the principle of necessity of self-identity, we can say that a is necessarily identical with a . From this it follows that it is true of a that it is necessarily identical with a . But from this and the assumption that a is identical with b it follows, by Leibniz's Law, that it is also true of b that it is necessarily identical with a – from which it follows that a is

necessarily identical with b . What we seem to have proved, then, is that if it is *true* that a is identical with b , then it is *necessarily true* that a is identical with b – and consequently that there cannot be an identity proposition that is merely *contingently* true (true in some possible worlds but not in others).

Lowe's more careful formulation of this argument is as follows (Lowe 2002: 85–86):

- (1) For any object x , it is necessarily the case that x is identical with x . [the necessity of self-identity]
 - (2) For any objects x and y , if x is identical with y , then whatever is true of x is also true of y . [Leibniz's Law]
 - (3) a is identical with b . [assumption]
 - (4) It is necessarily the case that a is identical with a . [from (1)]
 - (5) It is true of a that it is necessarily identical with a . [from (4)]
 - (6) If a is identical with b , then whatever is true of a is also true of b . [from (2)]
 - (7) Whatever is true of a is also true of b . [from (3) and (6)]
 - (8) It is true of b that it is necessarily identical with a . [from (5) and (7)]
 - (9) It is necessarily the case that a is identical with b . [from (8)]
- Therefore,
- (10) If a is identical with b , then it is necessarily the case that a is identical with b .
[from (3) & (9)]

Let us utilize lambda abstraction for highlighting the locution 'is true of': " $\lambda x(Fx)b$ ", or "It is true of b that it is an F ", for instance, is equivalent with " $\exists x(b=x \ \& \ Fx)$ ", but not, in general, with " Fb ", because the position of ' x ' in " Fx " may be intensional – for example, it may be

true of b that it is believed by c to be a G (or, $\lambda x(B_c Gx)b$), without it being true that c believes that b is a G (or, $B_c Gb$). On the face of it, Lowe's argument may be formalized as follows:

$$1^* \forall x \Box x = x$$

$$2^* \forall xy(x=y \rightarrow \forall X(\lambda z(Xz)x \leftrightarrow \lambda z(Xz)y))$$

$$3^* a=b$$

$$4^* \Box a=a$$

$$5^* \lambda z(\Box a=z)a$$

$$6^* a=b \rightarrow \forall X(\lambda z(Xz)a \leftrightarrow \lambda z(Xz)b)$$

$$7^* \forall X(\lambda z(Xz)a \leftrightarrow \lambda z(Xz)b)$$

$$8^* \lambda z(\Box a=z)b$$

$$9^* \Box a=b$$

$$10^* a=b \rightarrow \Box a=b$$

III. THE MISTAKE IN THE PUTATIVE ARGUMENT

Prima facie, the problem with Lowe's argument is that, apparently, the step from 4^* to 5^* works only if ' a ' is rigid – it is not true of the shortest spy that he or she is necessarily the shortest spy, or $\sim \lambda z(\Box s=z)s$ (even though it is true that $\Box s=s$) – and the step from 8^* to 9^* works only if ' b ' is rigid – it is not true that, necessarily, the number of apostles is twelve, or $\sim \Box t=n$ (even though it is true that $\lambda z(\Box t=z)n$). But positing that ' a ' and ' b ' are rigid trivializes the argument because then the truth of the conclusion 10^* is immediate (modulo the existence of a (i.e., b) in various possible worlds – though if we assume “Kaplan rigidity”, or direct referentiality, this existence proviso is not needed).

Lowe (2002: 89) notices this threat of trivialization. He says that he does not assume that ' a ' and ' b ' are rigid (because that would mean, according to Lowe (ibid.), that “there is

no possible world in which ‘*a*’ and ‘*b*’ designate different objects”) but only that they are “purely referential devices” (PRDs for short), which “serve purely to refer to certain objects and contribute nothing else to the proposition expressed with their aid” (Lowe 2002: 89; cf. Quine 1961: 140). In particular, definite descriptions are not allowed, for “genuine identity propositions” cannot, according to Lowe, even be formed by their means (Lowe 2002: 88).

However, PRDs – contributing, by definition, nothing but reference to the proposition – are directly referential terms (DRTs), for the standard characterization of the latter is that their semantic function is only to pick out the referent, or refer without a mediation of the content. And any DRT is rigid in a sense that may be clarified by means of *Carnapian intensions*, or (possibly incomplete) functions from possible worlds to entities (individuals, sets, or truth-values). The purpose of such intensions is to reflect the semantic roles of expressions. For a singular term ‘*b*’ in general, the Carnapian intension *I* (in the domain *A* of individuals) is a function from the set of possible worlds, *W*, to individuals ($I('b'): W \rightarrow A$), and for a *rigid* singular term ‘*c*’ it is a *constant* function ($I('c'): W \rightarrow \{c\}$). In contrast, for a singular term ‘*d*’ whose semantic function “is only to pick out the referent” (i.e., PRDs in Lowe’s terminology), the Carnapian intension – indicating the difference from rigidity – is not properly a function at all but reduces to singling out an individual ($I('d') = d$).³ Although a PRD may fail to be rigid in the strict technical sense of referring to the same individual *in* every possible world, we may still say that a PRD picks out the same individual *with respect to* (rather than *in*) every possible world, or that it is “Kaplan rigid”. All in all, if ‘*a*’ and ‘*b*’ are PRDs, and hence DRTs, on $a=b$ there is no possible world with respect to which ‘*a*’ and ‘*b*’ designate different objects. This trivializes Lowe’s argument.

Also, if ‘*a*’ and ‘*b*’ are PRDs – contributing to the proposition nothing else but the referent – then if ‘*a*’ and ‘*b*’ are co-referring, there cannot be any difference in the propositions *that a=a* and *that a=b*, which means that they are the same proposition (cf. the

next paragraph). Then, because it is true of *that* $a=a$ that it is necessary – $\lambda x(\Box x)p$, where p is *that* $a=a$ – this must, by PInI, be true of *that* $a=b$ as well: $\lambda x(\Box x)q$, where q is *that* $a=b$. This, again, trivializes the argument (for PInI is, of course, trivially true).

In addition to Lowe's argument being trivial, his resorting to PRDs has the following embarrassing consequence. Lowe seems to support the celebrated thesis that some necessarily true propositions, e.g. just "genuine identity propositions", are a posteriori. Indeed, he states that "if the argument [(1)–(10)] is correct, it shows us that there can be necessary truths that are not knowable a priori" (Lowe 2002: 86). However, as just indicated, if $a=b$, then *that* $a=a$ is the same proposition as *that* $a=b$, for PRDs ' a ' and ' b '. Thus, if it is true of *that* $a=a$ that it is a priori, this must, by PInI, be true of *that* $a=b$ as well.⁴ (Or perhaps both are a posteriori because the latter is – how are we to decide?) So, it seems that the proposition that $a=b$ is not a posteriori after all, a result that is unwelcome to almost everyone (though it has been argued for by some hardline direct reference theorists, for example Nathan Salmon (1986: 137–38)).

Lowe's mistake has to do with his misguided attention to signs. If we are trying to prove the necessity of identity of the object x with the object y , it cannot matter by which means $x (= y)$ is picked out – thus, definite descriptions, for example, should do as well. That is, it should not matter whether ' a ' and ' b ', as used in the proper argument for the necessity of identity, are rigid designators (e.g. proper names or indexicals) or nonrigid designators (e.g. definite descriptions). If the largest planet in the solar system (theL, for short) is identical with the planet which is fifth nearest to the sun (theF), then it is true of the planet that is in fact theL (= Jupiter) and the planet that is in fact theF (= Jupiter) that, necessarily, the former is identical with the latter.⁵ (By the same token, if the referent (or denotation or whatever term one wants to use) of 'theL' is the same as the referent of 'theF', then, necessarily, the former is identical with the latter.) The identity involved in the true "TheL is theF" is, like any identity, necessary identity, even though "TheL is theF" does not express a

necessary truth – there are possible worlds with respect to which “The largest planet is identical with the fifth nearest” is false, even though (to repeat) the objects that in fact are the largest planet and the fifth nearest planet are necessarily identical. It is just as Kripke (1980, 3) says:

If ‘*a*’ and ‘*b*’ are rigid designators, it follows that ‘*a=b*’, if true, is a necessary truth. If ‘*a*’ and ‘*b*’ are *not* rigid designators, no such conclusion follows about the *statement* ‘*a=b*’ (though the *objects* designated by ‘*a*’ and ‘*b*’ will be necessarily identical).

It might even be said that precisely the fact that definite descriptions won’t do in the putative argument shows that that argument, as Lowe presents it, cannot be used as an argument for the necessity of identity, and that this is a clear indication that those who formulate the argument in question in the manner Lowe does – as does, for instance, Bob Hale (2004: 365–67) – fail to understand it fully. One should not confuse the necessity of identity with the question of the necessary truth of identity statements.

Arthur Smullyan (1948) utilized the Russellian analysis of identity statements with definite descriptions as a way to dispel Barcan’s (1947) original version of (something like) the putative argument. Such a move is, I gather, behind the claim, made by Lowe (2002: 88) and others, that identity statements with definite descriptions are not “genuine identity statements”. However, such statements are, arguably, entirely legitimate identity statements (and Kripke seems to agree in the passage quoted above) simply because it is utterly natural to say things like “The largest planet in the solar system is the same as the planet which is fifth nearest to the sun” and infer from this and the fact that the volume of the largest planet is 1321 times the volume of the Earth that this is a fact about the fifth nearest planet as well. (To my mind, Lowe’s claim, already quoted above, that “there cannot be an identity proposition

that is merely contingently true” is misleading or at least unduly incautious.) In any case, we shall see shortly that at least as far as the argument for the necessity of identity is concerned, there is no need to resort to the “Smullyan Move” nor to quarrel over the correct “definition” of ‘identity statement’.

IV. THE NECESSITY OF IDENTITY

Let ‘*a*’ and ‘*b*’ be singular terms (any devices that pick out individuals), e.g., definite descriptions, proper names, or indexicals (as used in definite contexts). The true argument for the necessity of identity, independent of the nature of the singular terms used, goes as follows:

1' $\lambda xy(\Box x=y)aa$	necessity of self-identity
2' $a=b \rightarrow \forall X(\lambda x(Xx)a \leftrightarrow \lambda x(Xx)b)$	PI nI
3' $a=b$	assumption
4' $\lambda xy(\Box x=y)aa \rightarrow \lambda xy(\Box x=y)ab$	3' & 2' (for $\lambda x(Xx) = \lambda x(\lambda y(\Box x=y)a)$)
5' $\lambda xy(\Box x=y)ab$	1' & 4'
6' $a=b \rightarrow \lambda xy(\Box x=y)ab$	3' & 5', conclusion

The conclusion is that if *a* is identical with *b* then it is necessary that the former is identical with the latter. It does not matter whether ‘*a*’ and ‘*b*’, are rigid or nonrigid (or “purely referential” or not). For example, if the capital of Russia is the same as the most populous city in Europe, then it is true of the city that is in fact the capital of Russia and the city that is in fact the most populous city in Europe that, necessarily, the former (= Moscow) is identical with the latter (= Moscow). What is crucial for the conclusiveness of this proof of the necessity of identity (outdated Quinean qualms about the intelligibility of de re constructions

notwithstanding), is that ‘*a*’ and ‘*b*’ appear only in extensional positions – this means that the proof cannot be contested by appealing to intensionality.

V. THE ARGUMENT FROM SUBSTITUTIVITY

The general validity of PInI is sometimes doubted, usually on the basis that it (allegedly) fails for intensional statements (see Jacquette 2011 for a recent expression of the view that PInI is not universally applicable). It seems clear that this is a mistake: for example, if $a=b$ and it is true of a that it is believed by c to have the property F (or, $\lambda x(B_c Fx)a$), it is, obviously, true of b that it is believed by c to have the property F (or, $\lambda x(B_c Fx)b$).

Anyway, it can be seen from the argument given in the previous section that PInI is not even needed in the argument for the necessity of identity: it suffices to appeal to the hardly contestable Principle of Substitutivity in Extensional Positions (PSE): “If $a=b$ (so that ‘*a*’ and ‘*b*’ are co-referring) and ‘*a*’ occurs in a sentence in an extensional position, then ‘*a*’ may be replaced truth-preservingly by ‘*b*’ in that position in that sentence.”

The argument from substitutivity to the necessity of identity is short:

- | | |
|--|----------------------------|
| i. $a=b$ | assumption |
| ii. $\lambda xy(\Box x=y)aa$ | necessity of self-identity |
| iii. $\lambda xy(\Box x=y)ab$ | i & ii, by PSE |
| iv. $a=b \rightarrow \lambda xy(\Box x=y)ab$ | i & iii, conclusion |

For example, because the largest planet (= theL) is identical with the fifth planet (= theF) and it is true of the largest planet and the largest planet that necessarily, the former is identical with the latter, or $\lambda xy(\Box x=y)(theL)(theL)$,⁶ the latter ‘theL’ here, occurring in an extensional

position, may by PSE be replaced with ‘theF’, to yield the desired $\lambda xy(\Box x=y)(theL)(theF)$, or, “It is true of the largest planet and the fifth planet that, necessarily, the former is identical with the latter” (note again that I am not claiming that necessarily, the largest planet is identical with the fifth planet, or $\Box(theL = theF)$).⁷

ENDNOTES

¹ Of course, Barcan = Marcus (and even necessarily so).

² Lowe 2002 is a textbook, primarily aimed at students. So, is it an inappropriate object of criticism? No – to my mind it is particularly important to get things right in a textbook in order not to mislead students. In any case, similar mistakes appear in Lowe 2005: 300–4, and Lowe 2013: 128–33.

³ Cf. Kaplan 1989: 493, 497: “[A DRT is] an expression whose semantical rules provide directly that the referent in all possible circumstances [that is, possible worlds] is fixed to be the actual referent. [...] The referent, in a circumstance, of [...] a DRT] is simply independent of the circumstance and is [not] a function (constant or otherwise) of circumstance” (emphases removed). See also Maunu 2002: 55–60.

⁴ That is, $\lambda x(Rx)q$ follows by PInI from $p=q$ and $\lambda x(Rx)p$, where p is the proposition that $a=a$ and q the proposition that $a=b$ and ‘ R ’ stands for apriority. I think I need to emphasize here that ‘ p ’ is in $\lambda x(Rx)p$ in an *extensional* (de re) position, which means that the inference to $\lambda x(Rx)q$ is justified (given that $p=q$). There cannot be any threat of intensional fallacy, because the inference is *not* from Rp (where ‘ p ’ is in an intensional position) to Rq .

Some might protest that intensionality is not eliminated here because propositions themselves may be construed as *intensional objects*, or (I take it), as representable by (or as just *being*, according to some) functions from possible worlds to truth values (i.e., as

Carnapian intensions of declarative sentences). However, the argument stands even on this conception of propositions: If the function f is the same as the function g , then $\lambda x(Rx)f - 'f'$ not in an intensional position! – entails $\lambda x(Rx)g$, by PInI. (This issue was raised by an anonymous referee of *Journal of Philosophical Research*.)

⁵ Again, this is $theL = theF \rightarrow \lambda xy(\Box x=y)(theL)(theF)$. Definite descriptions ‘theL’ and ‘theF’ are in extensional positions in the consequent. The consequent is *not* the false $\Box(theL = theF)$.

⁶ The principle of the necessity of self-identity has sometimes been doubted on the basis that it is about *self*-identity of x rather than about the identity of x with x (see Lowe 2002: 86–87, Lowe 1982, McKay 1986). However, even though it is difficult to give a colloquial expression of “ $\lambda xy(\Box x=y)aa$ ” (perhaps somebody could be absent-minded enough to say, “The object a and the object a are such that necessary identity holds between them”), this cannot be held against the appropriateness of this statement: the unnaturalness of a statement when expressed in natural language has nothing to do with the correctness of the logical form of the statement.

⁷ I thank the anonymous referees of *Journal of Philosophical Research* for useful suggestions that improved this paper.

BIBLIOGRAPHY

Barcan, Ruth. 1947. “The Identity of Individuals in a Strict Functional Calculus of Second Order.” *Journal of Symbolic Logic* 12: 12–15.

Hale, Bob. 2004. “Putnam’s Retreat: Some Reflections on Hilary Putnam’s Changing Views about Metaphysical Necessity.” *Midwest Studies in Philosophy* 28: 351–378.

Jacquette, Dale. 2011. “Modal Objection to Naive Leibnizian Identity.” *History and*

Philosophy of Logic 32: 107–118.

Kaplan, David. 1989. “Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals.” In *Themes from Kaplan*, ed. John Almog, John Perry and Howard Wettstein. Oxford: Oxford University Press, 481–563.

Kripke, Saul. 1971. “Identity and Necessity.” In *Identity and Individuation*, ed. Milton K. Munitz. New York: New York University Press, 135–164.

-----, 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Lowe, E. J. 1982. “On the Alleged Necessity of True Identity Statements.” *Mind* 91: 579–584.

-----, 2002. *A Survey of Metaphysics*. Oxford: Oxford University Press.

-----, 2005. “Identity, Vagueness, and Modality.” In *Thought, Reference, and Experience: Themes from the Philosophy of Gareth Evans*, ed. José L. Bermúdez. Oxford: Oxford University Press, 290–310.

-----, 2013. *Forms of Thought: A Study in Philosophical Logic*. Cambridge: Cambridge University Press.

McKay, Tom. 1986. “Lowe and Baldwin on Modalities.” *Mind* 95: 499–505.

Marcus, Ruth Barcan. 1961. “Modalities and Intensional Languages.” *Synthese* 13: 303–322.

Maunu, Ari. 2002. “Natural Kind Terms Are Similar to Proper Names in Being World-Independent.” *Philosophical Writings* 19: 51–68.

Quine, Willard V. O. 1961. *From a Logical Point of View* (Second Revised Ed.). Cambridge: Harvard University Press.

Salmon, Nathan. 1986. *Frege’s Puzzle*, Atascadero: Ridgeview.

Smullyan, Arthur. 1948. “Modality and Description.” *Journal of Symbolic Logic* 13: 31–37.