Author version. Submitted to *The Oxford Handbook of Philosophy of Logic,*
edited by Brendel, Carrara, Ferrari, Hjortland, Sagi, and Sher.

The Necessity of Identity

Jessica Leech, King's College London.

1. Introduction

Identity is a peculiar notion. On the one hand, surely everything is just the thing that it is and nothing else. But on the other hand, we are tempted to think of it as a relation. After all, it appears as a relational predicate in formal languages (e.g. '$x = y$') and in natural languages (e.g. 'Eric is identical to George'). However, the idea of identity as a relation that might hold between two or more things is absurd for, after all, the whole idea of identity is that it concerns just one thing. At best, we can think of it as a relation that everything bears only to itself.

Nevertheless, the linguistic point is significant. In language—and in thought—identity appears as relational. This has consequences. While the very idea of one thing being the same thing as something else may seem absurd, the question of whether an identity statement in which the identity predicate is flanked by two different referring expressions is true is perfectly sensible, and has generated many pages of analytic philosophy. Further issues arise when we introduce modalities into the mix: whether true identity statements are necessarily or contingently so.

By 'identity' here I mean 'numerical identity', in the sense that if x is identical to y, then x and y are one and the same thing. This is a different notion to that of qualitative identity, where if x is qualitatively identical to y, then x and y have all and only the same qualities. It is widely assumed that (numerical) identity implies qualitative identity: if x is the same thing as y, then x and y have all and only the same qualities. This is often known as 'Leibniz's Law'.[1]

Leibniz's Law $\forall x \forall y(x = y \supset (Fx \equiv Fy))$

The reverse claim that qualitative identity implies (numerical) identity, *the identity of indiscernibles*, is more controversial.[2]

The aim of this chapter is to explore to some extent the relationship between identity and necessity in logic and metaphysics. First, I provide a historically-based summary of proofs of the necessity of identity, highlighting the importance of the role that *self-identity* plays. Second, I introduce two examples of metaphysical topics where the necessity of identity has played a pivotal role: the necessary *a posteriori*, and the coincidence of material objects. I argue that important aspects of these debates rest on how we *represent* identity. Third, I consider some recent work on *generalized identity*. This opens up new prospects for explaining *why* identity is necessary.

A brief word on notation. There are many conditionals in this chapter. I normally use '⊃' to signify the material conditional and '≡' to signify the material biconditional. Where it is appropriate to follow older conventions, '⥽' signifies the

[1] This statement of Leibniz's Law is a schema to be instantiated by any instance of the predicate placeholder F.

[2] See Forrest (2020).

strict conditional and '≣' the strict biconditional.[3] In the final section, following recent convention, I use '≡' to signify generalized identity, and so '↔' for the material biconditional there.

2. Identity and Necessity in Formal Logic

The aim of this section is to look at formal proofs of the necessity of identity. As we shall see, how such a proof goes, and whether it is successful, depends upon several choices to be made in one's formal logical system. The first task is to present several proofs of the necessity of identity: from Barcan, Quine, and Kripke. Part of my aim here is to spell out Barcan's proofs, in order to clarify and emphasize Barcan's original achievement. I then bring out some key similarities and differences of the various proofs, placing emphasis on the importance of self-identity.

The proof of the necessity of identity begins with Ruth Barcan.[4] Her proof appears in a technical paper of 1947, rich with dense and unfamiliar formalism. The more familiar version of the proof was presented and popularized by Saul Kripke. Attribution of the proof to Barcan has been contested, in particular by Soames (1995) and Burgess (2014) who contend that, while Barcan certainly proved some important results in the vicinity, her paper does *not* contain the crucial version of the proof as we recognise it today. Barcan's proof, it is argued, is markedly different and depends

[3] Barcan (1946a, 2): $(A \equiv B) =_{df} ((A \prec B) \cdot (B \prec A))$. There is a typo in the Barcan paper: both material and strict equivalence are introduced as a triple bar, when of course the latter should be the quadruple bar.

[4] See, e.g., mentions in Della Rocca (1996, 188), Fitting and Muskens (2022), Hale (2013, 260), Williamson (2013, 26).

upon controversial modal assumptions, most importantly, the 4 axiom distinctive of S4 systems, and second-order versions of the Barcan formulas.[5] In fact, suggests Burgess (2014), the first sketch of the familiar proof is in Quine (1953). Arguably, Arthur Prior also got there around the same time (Kürbis, forthcoming).

I here defend Barcan's proof to an extent. I grant that the 1947 proof is not, indeed, the version of the proof to be found in Quine's and Kripke's work. That is plain enough from the fact that Barcan is working in a different logical system. Barcan's system is second-order (the Quine-Kripke proof is ostensibly first-order); Barcan defines identity, rather than taking it to be primitive; and Barcan's system does not include a necessitation rule. I will argue, however, that the crucial moves of the familiar proof are clearly present in Barcan's version: a law of substitutivity, self-identity, and theorems and rules governing conditionals.

I will present three different proofs. First, the familiar "Quine-Kripke" proof. Second, Barcan's proof of the *material equivalence* of two definitions of identity: what I shall call "material identity" and "strict identity". Third, Barcan's proof of the *strict equivalence* of material and strict identity.

2.1. The Quine-Kripke proof

[5] Barcan Formula (first-order): $\forall x \Box \Phi x \supset \Box \forall x \Phi x$. Converse Barcan Formula (first-order): $\Box \forall x \Phi x \supset \forall x \Box \Phi x$. Second-order Barcan: $\forall X \Box \Phi \supset \Box \forall X \Phi$. Second-order Converse Barcan: $\Box \forall X \Phi \supset \forall X \Box \Phi$. I have presented these as material conditionals. If they are theorems of a system with a necessitation rule, the strict conditionals also hold. Barcan's own versions directly state the strict conditionals as theorems.

The proof of the necessity of identity is probably most familiar from Kripke's presentation.

> First, the law of substitutivity of identity says that, for any objects x and y, if x is identical to y, then if x has a certain property F, so does y:
>
> (1) $(x)(y)[(x = y) \supset (Fx \supset Fy)]$
>
> On the other hand, every object surely is necessarily self-identical:
>
> (2) $(x)\Box(x = x)$
>
> But
>
> (3) $(x)(y)(x = y) \supset [\Box(x = x) \supset \Box(x = y)]$
>
> is a substitution instance of (1), the substitutivity law. From (2) and (3), we can conclude that, for every x and y, if x equals y, then, it is necessary that x equals y:
>
> (4) $(x)(y)((x = y) \supset \Box(x = y))$
>
> This is because the clause $\Box(x = x)$ of the conditional drops out because it is known to be true. (Kripke, 1971, 136)

This proof relies on three principles. First, a law of substitutivity of identity (Leibniz's Law). Second, the necessity of self-identity. Third, a principle governing conditionals: if $A \supset (B \supset C)$ and B then $(A \supset C)$.[6]

---

[6] Kripke narrates the proof in an unfortunately epistemic way. Of course, it doesn't matter for a proof of the necessity of identity whether the necessity of self-identity is *known*. But if it is known, it is true. So I here ignore the epistemic tone and simply take the important move to be that (2) is true.

One can find the proof presented in this way earlier, in Quine (1953).[7]

> There is a more fundamental form of the law of substitutivity of identity ... viz.:
>
> (51) $(x)(y)(x = y. \supset . Fx \equiv Fy)$
>
> ... The generality of 'F' in (51) is this: 'Fx' is to be interpretable as any open sentence of the system in question, having 'x' as free (quantifiable) variable ... If 'nec' is not referentially opaque, 'Fx' and 'Fy' in (51) can in particular be taken respectively as 'nec$(x = x)$' and 'nec$(x = y)$'. From (51), therefore, since surely 'nec$(x = x)$' is true for all x, we have:
>
> (52) $(x)(y)[x = y. \supset \text{nec}(x = y)]$.
>
> I.e., identity holds necessarily if it holds at all. (Quine, 1953, 173)

The proof relies on the same three steps: substitutivity of identity; necessity of self-identity; and the logic of conditionals. I.e., spelling out the latter steps in Quine's argument, an instance of (51) is

$$(x)(y)(x = y. \supset . (\text{nec}(x = x) \supset \text{nec}(x = y))).$$

It is true that nec$(x = x)$. So, since if $A \supset (B \supset C)$ and B then $(A \supset C)$, we may infer (52).

A proof also appears in Arthur Prior's *Formal Logic*. Prior presents his proof as 'due in substance' to Barcan (1947) (Prior, 1962, 205, n.1). For reasons of space, I will confine my comparisons to Quine, Kripke and Barcan, but it is important to note

[7] See also Wiggins (1965).

Prior's contribution. Kürbis (forthcoming) suggests that Prior developed this version of the proof independently from and around the same time as Quine.[8]

2.2. The material equivalence of material and strict identity in S2[2]

Barcan's first proof is carried out in S2[2], a second-order system of S2. For details of this system see Barcan (1946a,b, 1947) and Lewis and Langford (1932) Appendix II.[9] The important details for present purposes are that this system does not include a necessitation rule, *if* $\vdash A$ *then* $\vdash \Box A$, but it does include the T-axiom: $\Box p \prec p$.[10]

---

[8] 'Prior's proof is on p.205f of the second edition of *Formal Logic*. This edition, says the preface to it, left the substance of the text of the first edition untouched. In a footnote to p. 205, Prior attributes his proof as 'in substance' due to Barcan. He does not mention Quine. This indicates that Prior was not aware of Quine's proof at the time of writing *Formal Logic*. Prior was a prolific referencer, so had he known Quine's proof, we can be fairly confident that he would have said so. Another reason is that he probably could not have seen Quine's proof in print at the time of the completion of *Formal Logic*. Although the first edition of *Formal Logic* was published only in 1955, the preface is dated to May 1953. Assuming the preface was written last and any changes to the body of the text afterwards restricted to minor corrections at proof stage, we can conclude with some confidence that Quine's publications were not available to Prior at the time and may not yet have been available in print at all.' (Kürbis, forthcoming, 2)

[9] See also Ballarin (2021).

[10] Lewis and Langford specify that S2 includes 'all of the theorems of section 1–5 in Chapter VI' (1932, 500). Theorem 18.42, that 'what is necessary is true' (1932, 163),

In her second-order systems, Barcan proposes two different ways to define identity. "Strict identity" is defined in terms of Leibniz's Law expressed using a strict conditional. In Barcan's symbolism:

$$I =_{df} \hat{\alpha}_1\hat{\alpha}_2(\theta)(\theta(\alpha_1) \prec \theta(\alpha_2))$$

This defines the relation I. It says that I is that relation which holds between α_1 and α_2 just when necessarily α_1 and α_2 have all any only the same attributes. Barcan uses abstraction operators. Using more familiar notation and λ-abstraction, we can present this as,

$$I =_{df} \lambda x \lambda y \forall X \Box (Xx \supset Xy)$$

This definition implies that x and y are strictly identical just when, necessarily, x and y have all and only the same attributes.

$$xIy \equiv \forall X \Box (Xx \supset Xy)$$

"Material identity" is then defined in terms of a material conditional.

---

appears in Chapter VI section 4. Lewis and Langford use '$\neg\Diamond\neg p$' plus primitive strict implication. I will sometimes use '$\Box p$' for '$\neg\Diamond\neg p$' and '$\Box(p \supset q)$' for '$p \prec q$'. See Barcan (1946a, 2). Here it is important to specify the strict conditional, since S2[2] lacks a necessitation rule, and so $\vdash \Box p \supset p$ does not guarantee alone that $\Box(\Box p \supset p)$.

$$I_m =_{df} \hat{\alpha}_1\hat{\alpha}_2(\theta)(\theta(\alpha_1) \supset \theta(\alpha_2))$$

This says that I_m is that relation which holds between α_1 and α_2 just when α_1 and α_2 have all any only the same attributes. Updating again, we have:

$$I_m =_{df} \lambda x \lambda y \forall X(Xx \supset Xy)$$

This implies that x and y are materially identical just when they have all and only the same attributes:

$$xI_m y \equiv \forall X(Xx \supset Xy)$$

For simplicity, in what follows I will drop the abstraction operators in the updated notation.[11]

[11] As do Burgess (2014) and Soames (1995). Williamson (2013) uses λ-abstraction, noting that Barcan's use of an abstraction operator 'permits a contingentist to accept the being constraint for strict and material identity' Williamson (2013, 203, n.10). The being constraint 'says that being something is a necessary condition for having properties or relations'. In our case, it allows one to accept that xIy is true only if x and y exist. However, 'by contrast, if Ixy and $I_m xy$ simply abbreviate $\forall X\Box(Xx \to Xy)$ and $\forall X(Xx \to Xy)$ respectively, then Ixx and $I_m xx$ abbreviate $\forall X\Box(Xx \to Xx)$ and $\forall X(Xx \to Xx)$, which hold independently of $\exists y x = y$' (Williamson, 2013, 203, n.10).

Theorem 2.31 of Barcan (1947) is the material equivalence of material and strict identity.[12]

2.31 $\vdash (\beta_1 I_m \beta_2) \equiv (\beta_1 I \beta_2)$

[2.31] $\vdash \forall X(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy)$

I'll first present Barcan's proof, explain it in her terms, then present a version in more familiar terms.[13]

1.	$(\beta_1 I_m \beta_2) \supset ((\beta_1 I \beta_1) \supset (\beta_1 I \beta_2))$	2.21, *14.1*, mod pon, 2.3, subst
2.	$(\beta_1 I_m \beta_2) \supset (\beta_1 I \beta_2))$	*15.8*, subst, 2.6, adj, *14.29*, mod pon
3.	$(\beta_1 I \beta_2) \supset (\beta_1 I_m \beta_2))$	*12.1*, 2.23, subst, *18.42*, VIII, *14.1*, mod pon
4.	$(\beta_1 I_m \beta_2) \equiv (\beta_1 I \beta_2))$	adj, def

Barcan makes us do our homework. Whilst theorems referred to by numbers beginning with "2" appear earlier in the same paper (Barcan, 1947), italicised numbers refer to theorems in Lewis and Langford's *Symbolic Logic* (Lewis and

[12] I use square brackets to signify the translation of a theorem into familiar notation without abstraction.

[13] In Barcan (1947, 15) the proof contains a typo in the main result, showing a quadruple bar for strict equivalence, '≣', rather than '≡' for material equivalence. See Fitch (1949), footnote 3, for corrections of this and other typographical errors.

Langford, 1932), and further theorems and rules—numbers beginning with "1" and roman numerals—are taken from Barcan's earlier papers (1946a; 1946b).

2.21 is a principle of substitution: necessarily, if β_1 is materially identical to β_2, then, for all properties, if β_1 has a property, β_2 has it too.[14]

2.21 $\vdash (\beta_1 I_m \beta_2) \prec ((\beta_1 \in \hat{\alpha}A) \supset (\beta_2 \in \hat{\alpha}A))$

2.3 is an abstraction principle: the proposition that some things have a property A-ness is strictly equivalent to the proposition that those things A, i.e., "x has the property F-ness" is strictly equivalent to "x is F".

2.3 $\hat{\alpha}_1\hat{\alpha}_2 \ldots \hat{\alpha}_n A(\beta_1, \beta_2, \ldots, \beta_n) \equiv B$ where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are distinct individual variables occurring freely in A, no free occurrence of α_m $(1 \leq m \leq n)$ in A is in a wf part of A of the form $(\beta_m)\Gamma$, and B results from A by replacing all free occurrences of α_1 by β_1, all free occurrences of α_2 by β_2, ... , all free occurrences of α_n by β_n in A.

14.1 is a principle governing conditionals: necessarily, if p strictly-implies q, then p materially-implies q:

14.1 $(p \prec q) \prec (p \supset q)$.

[14] Here and throughout we need not take such property talk seriously. One can think of property talk in terms of abstraction locutions based on predications.

We are now in a position to spell out step one of the proof. According to 2.21, necessarily, if β_1 is materially identical to β_2, then, for all properties, if β_1 has a property, β_2 has it too. By *14.1* and *modus ponens*, we move from the strict to the material conditional: if β_1 is materially identical to β_1, then, for all properties, if β_1 has a property, β_2 has it too. An instance of this concerns *being strictly identical to* β_1, i.e.,

$$(\beta_1 I_m \beta_2) \supset ((\beta_1 \in \hat{\alpha}(\beta_1 I \alpha)) \supset (\beta_2 \in \hat{\alpha}(\beta_1 I \alpha)))$$

By 2.3 and substitution, we move from the abstraction to the predication.

$$(\beta_1 I_m \beta_2) \supset ((\beta_1 I \beta_1) \supset (\beta_1 I \beta_2))$$

The second step of the proof condenses the conditional via the introduction of self-strict-identity as a theorem:

2.6 $\vdash \beta I \beta$

Barcan appeals to two further theorems concerning conditionals:

15.8 $\big((p \wedge q) \supset r\big) \coloneqq \big(p \supset (q \supset r)\big) \coloneqq (q \supset (p \supset r))$

14.29 $(p \wedge (p \supset q)) \dashv q$

So, we have it that $(\beta_1 I_m \beta_2) \supset \big((\beta_1 I \beta_1) \supset (\beta_1 I \beta_2)\big)$. By *15.8* and substitution,[15] this gives us $(\beta_1 I \beta_1) \supset ((\beta_1 I_m \beta_2) \supset \big((\beta_1 I \beta_2)\big)$. But since $\beta_1 I \beta_1$ (2.6), we can conjoin this,[16] instantiate *14.29*,[17] and apply *modus ponens* (for the strict conditional), to give us $(\beta_1 I_m \beta_2) \supset (\beta_1 I \beta_2)$. This provides the left-to-right direction of the material equivalence.

It remains to establish the right-to-left direction. *12.1* is the theorem that everything strictly implies itself: $\Box(p \supset p)$. In particular, $\Box((\beta_1 I \beta_2) \supset (\beta_1 I \beta_2))$. Theorem 2.23 states the strict equivalence of strict identity and necessary material identity.

2.23 $\vdash \Box(\beta_1 I_m \beta_2) \equiv (\beta_1 I \beta_2)$

This allows us to substitute to give $\Box((\beta_1 I \beta_2) \supset \Box(\beta_1 I_m \beta_2))$. An instance of *18.42*, the T axiom, is $\Box(\Box(\beta_1 I_m \beta_2) \supset (\beta_1 I_m \beta_2))$. By transitivity for strict implication, VIII,[18] we thus have $\Box((\beta_1 I \beta_2) \supset (\beta_1 I_m \beta_2))$. An instance of *14.1* is $\Box(\Box((\beta_1 I \beta_2) \supset (\beta_1 I_m \beta_2)) \supset ((\beta_1 I \beta_2) \supset (\beta_1 I_m \beta_2))$. So via *modus ponens* we can infer $(\beta_1 I \beta_2) \supset (\beta_1 I_m \beta_2)$.

Let us summarize in more familiar terms. Since I am not using abstraction operators, I won't need to explicitly employ 2.3. Strictly speaking, first-order variables (x, y) should be bound by universal quantifiers, but I omit those here for readability.

---

[15] Where p is $\beta_1 I_m \beta_2$, q is $\beta_1 I \beta_1$, and r is $\beta_1 I \beta_2$.

[16] $(\beta_1 I \beta_1) \wedge ((\beta_1 I \beta_1) \supset \big((\beta_1 I_m \beta_2) \supset (\beta_1 I \beta_2)\big))$

[17] $(\beta_1 I \beta_1) \wedge ((\beta_1 I \beta_1) \supset \big((\beta_1 I_m \beta_2) \supset (\beta_1 I \beta_2)\big)) \prec ((\beta_1 I_m \beta_2) \supset (\beta_1 I \beta_2))$

[18] If $\vdash A_1 \prec A_2, \vdash A_2 \prec A_3, \ldots, \vdash A_{n-1} \prec A_n$ then $\vdash A_1 \prec A_n$.

1.	$\Box(\forall X(Xx \supset Xy) \supset (\forall X\Box(Xx \supset Xx) \supset \forall X\Box(Xx \supset Xy)))$	Instance of 2.21
2.	$\forall X(Xx \supset Xy) \supset (\forall X\Box(Xx \supset Xx) \supset \forall X\Box(Xx \supset Xy)))$	1,*14.1*, MP
3.	$\forall X\Box(Xx \supset Xx) \supset (\forall X(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy)))$	2, *15.8,* substitution
4.	$\forall X\Box(Xx \supset Xx)$	Instance of 2.6
5.	$\forall X\Box(Xx \supset Xx) \wedge (\forall X\Box(Xx \supset Xx)$ $\supset (\forall X(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy))))$	3,4, adj.
6.	$\Box(\forall X\Box(Xx \supset Xx) \wedge (\forall X\Box(Xx \supset Xx)$ $\supset (\forall X(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy)))))$ $\supset (\forall X(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy)))$	Instance of *14.29*
7.	$\forall X(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy)$	5,6, MP
8.	$\Box(\forall X\Box(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy))$	Instance of *12.1*
9.	$\Box\forall X(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy)$	2.23
10.	$\Box(\forall X\Box(Xx \supset Xy) \supset \Box\forall X(Xx \supset Xy))$	8,9, substitution
11.	$\Box(\Box\forall X(Xx \supset Xy) \supset \forall X(Xx \supset Xy))$	Instance of T
12.	$\Box(\forall X\Box(Xx \supset Xy) \supset \forall X(Xx \supset Xy))$	10,11,VIII
13.	$\Box(\Box(\forall X\Box(Xx \supset Xy) \supset \forall X(Xx \supset Xy)))$ $\supset (\forall X\Box(Xx \supset Xy) \supset \forall X(Xx \supset Xy)))$	Instance of *14.1*
14.	$\forall X\Box(Xx \supset Xy) \supset \forall X(Xx \supset Xy)$	12,13, MP
15.	$\forall X(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy)$	7,14, ≡-introduction

The proof begins with a principle of substitution. If $\forall X(Xx \supset Xy)$, then, if a predicate instance of X is '$\forall X\Box(Xx \supset X \ldots)$', we can substitute in y, i.e., $\forall X\Box(Xx \supset Xy)$. This, together with theorems and rules governing the conditionals, plus the theorem of self-strict-identity, yields the result, at step 7, that material identity materially-implies strict identity.

The second half of the proof depends upon the introduction of theorem 2.23 (step 9). That theorem, when cashed out in terms of the definitions of identity, is an application of the Barcan Formulas at second order, i.e., $\Box\forall X\Phi \equiv \forall X\Box\Phi$. Via rules and theorems governing the conditionals, along with the T axiom, we yield the result that strict identity materially-implies material identity. In the final step, these two results at 7 and 14 combine to give the material equivalence of material and strict identity.

2.3. The strict equivalence of material and strict identity in S4^2

Theorem 2.33* of Barcan (1947) states the strict equivalence of material and strict identity. The system is second order S4: formulas are starred when they are theorems of S4^2 that go beyond S2^2. Again, for details on the system see Barcan's papers and Lewis and Langford (1932).[19] The important addition for us here is Lewis and Langford's C10: $\neg\Diamond\neg p \prec \neg\Diamond\neg\neg\Diamond\neg p$ (equivalently $\Box p \prec \Box\Box p$).[20] In a system which already contains *18.42* (the T axiom), they note that a second form of C10 can be derived: $\Box p = \Box\Box p$. This is included in Barcan (1946b) as 104, and in Barcan (1947) as 1.104. I shall again present Barcan's proof, then summarize in more familiar terms.

The key result is:

2.33* $\vdash (\beta_1 I_m \beta_2) \equiv (\beta_1 I \beta_2)$

[19] See also Ballarin (2021).

[20] Lewis and Langford (1932, 497).

[2.33*] $\vdash \Box(\forall X(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy))$

Barcan's proof is:

1.	$((\beta_1 I_m \beta_2) \wedge (\beta_1 I \beta_1)) \prec (\beta_1 I \beta_2)$	2.21, 2.3, subst, *14.26*
2.	$(\beta_1 I_m \beta_2) \prec (\beta_1 I \beta_2)$	2.6, 2.32*, subst, adj, *18.61*, mod pon
3.	$(\beta_1 I \beta_2) \equiv (\beta_1 I_m \beta_2)$	*18.42*, 2.23, subst, adj, def

The first step is similar to the proof of material equivalence, but it appeals to a different theorem about conditionals:

14.26 $(p \wedge q) \prec r \coloneqq p \prec (q \supset r) \coloneqq q \prec (p \supset r)$

An instance of the substitution principle 2.21, plus abstraction principle 2.3, gives us $(\beta_1 I_m \beta_2) \prec ((\beta_1 I \beta_1) \supset (\beta_1 I \beta_2))$. Given *14.26*, we can substitute this for $((\beta_1 I_m \beta_2) \wedge (\beta_1 I \beta_1)) \prec (\beta_1 I \beta_2)$.[21]

Step two packs in a great deal of action. The key theorem appealed to is 2.32*, which is proved in Barcan's paper just prior to 2.33*. 2.32* states the strict equivalence of strict identity and necessary strict identity.[22]

2.32* $\vdash \Box(\beta_1 I \beta_2) \equiv (\beta_1 I \beta_2)$ [2.23, 1.104, subst]

[21] Where p is $\beta_1 I_m \beta_2$, q is $\beta_1 I \beta_1$, and r is $\beta_1 I \beta_2$.

[22] Barcan (1947) contains another typo here, missing out the equivalence connective entirely. See Fitch (1949), footnote 3.

[2.32*] $\vdash \Box(\Box\forall X\Box(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy))$

Barcan proves this by appeal to 2.23, which we saw is effectively an application of second order Barcan Formulas, plus 1.104.

1.104 $\vdash \Box\Box A \equiv \Box A$

With these two theorems, the proof of 2.32* is fairly straightforward.[23] An instance of 1.104 is $\Box\Box(\beta_1 I_m \beta_2) \equiv \Box(\beta_1 I_m \beta_2)$. Given 2.23, we can substitute both instances of $\Box(\beta_1 I_m \beta_2)$ for $(\beta_1 I \beta_2)$, which gives us $\Box(\beta_1 I \beta_2) \equiv (\beta_1 I \beta_2)$.

This, then, is how step two of the proof of 2.33* goes. We have $((\beta_1 I_m \beta_2) \wedge (\beta_1 I \beta_1)) \prec (\beta_1 I \beta_2)$. It is a theorem that $\beta_1 I \beta_1$. Given 2.32*, we can substitute $\beta_1 I \beta_1$ for $\Box(\beta_1 I \beta_1)$. We conjoin to give $\Box(\beta_1 I \beta_1) \wedge (((\beta_1 I_m \beta_2) \wedge (\beta_1 I \beta_1)) \prec (\beta_1 I \beta_2))$. Given theorem *18.61*—

18.61 $(\Box p \wedge ((p \wedge q) \prec r)) \prec (q \prec r)$

—and *modus ponens*, we may infer $(\beta_1 I_m \beta_2) \prec (\beta_1 I \beta_2)$: necessarily, if some things are materially identical, then they are strictly identical.

The other direction is then straightforward. An instance of (one direction of) 2.23 is $(\beta_1 I \beta_2) \prec \Box(\beta_1 I_m \beta_2)$. Given T, it follows that $(\beta_1 I \beta_2) \prec (\beta_1 I_m \beta_2)$.

[23] Thank you to Julien Dutant for the presentation of this proof.

In sum:[24]

1.	$\Box(\forall X(Xx \supset Xy) \supset (\forall X\Box(Xx \supset Xx) \supset \forall X\Box(Xx \supset Xy)))$	Instance of 2.21
2.	$\Box((\forall X(Xx \supset Xy) \wedge \forall X\Box(Xx \supset Xx)) \supset \forall X\Box(Xx \supset Xy))$	1,*14.26*, substitution
3.	$\forall X\Box(Xx \supset Xx)$	Instance of 2.6
4.	$\Box\Box\forall X(Xx \supset Xy) \equiv \Box\forall X(Xx \supset Xy)$	1.104
5.	$\Box\forall X(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy)$	2.23
6.	$\Box\forall X\Box(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy)$	4,5, substitution
7.	$\Box\forall X\Box(Xx \supset Xx)$	3,6, substitution
8.	$\Box\forall X\Box(Xx \supset Xx) \wedge \Box((\forall X(Xx \supset Xy) \wedge \forall X\Box(Xx \supset Xx)) \supset \forall X\Box(Xx \supset Xy))$	2,7, adj.
9.	$\Box(\Box\forall X\Box(Xx \supset Xx) \wedge \Box((\forall X(Xx \supset Xy) \wedge \forall X\Box(Xx \supset Xx)) \supset \forall X\Box(Xx \supset Xy))) \supset \Box(\forall X(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy)))$	Instance of *18.61*
10.	$\Box(\forall X(Xx \supset Xy) \supset \forall X\Box(Xx \supset Xy))$	8,9, MP
11.	$\Box(\forall X\Box(Xx \supset Xy) \supset \Box\forall X(Xx \supset Xy))$	Instance of 2.23
12.	$\Box(\forall X\Box(Xx \supset Xy) \supset \forall X(Xx \supset Xy))$	11, T, VIII
15.	$\forall X(Xx \supset Xy) \equiv \forall X\Box(Xx \supset Xy)$	10,12, $\equiv$-introduction

2.4. Remarks on these proofs

[24] Again, omitting first-order quantifiers for readability.

The first point to emphasise is that the three proofs just presented share three key features: (1) a law of substitutivity; (2) self-identity; (3) contraction of a conditional.

All three proofs begin with a law of substitutivity (principle of substitution), and an instance of that law where the predicate in question concerns identity with something. Such a law of substitutivity, which allows for substitution of complex modal predicates such as '$\Box(x = \ldots)$', is particularly permissive, and for that reason potentially suspect. For example, Ben-Yami (2018), argues that an extension of such a law beyond substitution of atomic formulas is not justified. However, we shall set this kind of concern to one side for present purposes.[25]

The next important step involves self-identity. The Quine-Kripke proof introduces the *necessity* of self-identity ($(x)\Box(x = x)$). The crucial step in the Barcan proofs is the *theorem* of self-strict-identity ($\vdash \beta I \beta$). In the second Barcan proof, this contributes to a derivation of necessary self-strict-identity ($\Box\beta I \beta$). These three very similar steps in the proofs play the same role; providing that central antecedent of an instance of the law of substitutivity, which is via further steps to be removed. In the proof of material equivalence, theorem 2.6 is sufficient for application of theorems for the conditional to yield one direction: $(\beta_1 I_m \beta_2) \supset (\beta_1 I \beta_2)$. In the proof of strict equivalence things are less straightforward. The crucial theorem for the conditional, *18.61*, applies here just if self-strict-identity is necessary, i.e. $\Box\beta I \beta$. As Fitch (1949), footnote 3, notes, if Barcan was working in a system with the necessitation rule, she

[25] See also Burgess (2014) for further discussion on the success and significance of such derivations.

could infer this necessity from the theorem of self-strict-identity.[26] Since Barcan does not have that rule available, the next steps of her proof are more complicated. She appeals to 2.32*: $\Box(\beta_1 I \beta_2) \equiv (\beta_1 I \beta_2)$. With that, she can substitute $\beta I \beta$ for $\Box\beta I \beta$. However, as we saw earlier, to show that $\beta_1 I \beta_2$ strictly-implies $\Box(\beta_1 I \beta_2)$, Barcan uses the 4 axiom.

Does this difference have any particular significance, beyond technical interest? I believe so. It is an important question what kind of necessity is implicated in the necessity of identity. Adoption of a necessitation rule builds in an assumption about the kind of necessity, concerning its relationship with theoremhood: *if something is a theorem, then it is necessary*. It may seem plausible that such a relationship should hold, but in developing a modal logic from the ground up, and in considering basic questions of how notions such as identity or quantification may interact with modalities, it is nevertheless a significant assumption. One can see why someone such as Barcan, working at the early frontiers of modal logics, might choose to avoid such an assumption. Instead, she appeals to assumptions concerning modal operators themselves (such as the 4 axiom) and their interaction with quantifiers (the Barcan formulas), independently of the modal status of theorems. Without further consideration of what kind of modality we aim to capture here, it is not at all clear which set of assumptions is preferable. We will return to this question in a moment.

The final step of the Quine-Kripke proof, and the steps of the Barcan proofs that give us left-to-right, concern theorems and rules for conditionals that effectively

[26] This is how Prior proceeds explicitly at step 6 of his proof, since he has a necessitation rule (Prior, 1962, 205f., Kürbis, forthcoming, 3).

allow us to contract our instance of the law of substitutivity, by taking out the central statement of (the necessity of) self-identity.

In sum, are these three proofs different? *Yes, of course*. Do they follow the same three basic moves? *Also yes*. Why are they different? Not least because Barcan is using different logical systems. But the *overall shape* of proof is the same.

The question remains: are any of Barcan's proofs to be counted as proofs *of the necessity of identity*? The proofs of material (2.31) and strict equivalence (2.33*) both resemble the Quine-Kripke proof in key respects. However, some readings of Barcan suggest that we should instead take 2.32* to be the necessity of identity, i.e., $\Box(\beta_1 I \beta_2) \equiv (\beta_1 I \beta_2)$.[27] Indeed, Barcan herself later seems to say this.

> The following are theorems of QS4:[28]
>
> (7) $(xI_m y) \equiv (xIy)$
>
> (8) $(xIy) \equiv \Box(xIy)$
>
> where '$\Box$' is the modal symbol for logical necessity. In (7) 'I_m' and 'I' are strictly equivalent; within such a modal language, they are therefore indistinguishable by virtue of the substitution theorem. Contingent identities are disallowed by (8). (Barcan Marcus, 1961, 9)

(7) here is 2.33* and (8) is 2.32*. If indeed 2.32* is the statement of the necessity of identity, then Soames (1995) and Burgess (2014) are correct in their claims that this

[27] See Smith (1995, 184), Soames (1995, 209, note 4).

[28] Barcan's system QS4 adds the Barcan Formulas to Lewis's S4.

is nothing like the Quine-Kripke proof. As we saw, the proof of 2.32* traded just on the Barcan formulas (2.23) and the 4 axiom (1.104).

Both strict identity and necessary strict identity are candidates for identity necessitated, and so both 2.32* and 2.33* are candidate statements of the necessity of identity. Is one of them to be rightly called “*the* necessity of identity”. Perhaps one could argue one way or another. However, as should already be clear, neither of these are strictly the same as the conclusion of the Quine-Kripke proof, since we are working with a different language in which identity is defined. Barcan was exploring different ways to formalise the notion of identity. Ultimately, once all of her results are put together, we see that the different ways to necessitate identity prove to be equivalent.

$$(xIy) \equiv \Box(xIy) \equiv \Box(xI_m y)$$

$$\forall X\Box(Xx \supset Xy) \equiv \Box\forall X\Box(Xx \supset Xy) \equiv \Box\forall X(Xx \supset Xy)$$

In which case, it is hardly in the spirit of the work to declare that just one result is *the* necessity of identity. Indeed, we saw Barcan point out that ‘within such a modal language’ the different definitions of necessity are ‘indistinguishable’. There is a broader picture here, to which all of her proofs add important elements. But, evidently, the necessity of identity is present in Barcan (1947), as is the shape of the familiar Quine-Kripke proof of it.

Barcan’s proof requires a stronger modal system than the Quine-Kripke version. But this is precisely because they are playing by very different rules: Kripke and Quine *don’t* provide definitions of identity and *do* assume the necessity of self-identity. As we have seen, Barcan also isolates assumptions about the modalities to the behaviour of modal operators (including the strict conditional) *within* the logic, eschewing a necessitation rule which ties the modalities to a metalogical property of theoremhood. The moral to be drawn here is that how, if at all, one can prove the necessity of identity varies according to many crucial assumptions, including substantive questions concerning whether or not, and if so how, a formal system should seek to define an identity predicate, and what (kind of) principles should govern its modal terms.

Let us return to the question: with what kind of necessity are these proofs of the necessity of identity concerned? The obvious candidates are logical necessity and metaphysical necessity. There is much to be said about these, but let us assume two key ideas: (1) logical necessity is closely related to logical validity, and so is sensitive to linguistic differences; (2) metaphysical necessity concerns things as they are independently of the linguistic means used to talk about them. We saw in Quine's proof that he explicitly states the assumption that "'nec' is not referentially opaque", suggesting that the necessity here is metaphysical not logical. We have seen that Barcan's proof, by not assuming a necessitation rule, and so not assuming a close relationship between theoremhood and necessity, also puts distance between logical necessity and the necessity at issue. Kripke famously defends a distinctively metaphysical notion of necessity that can accommodate essentialist claims, in contrast to notions of analyticity and *a priority* (see Section 3.1). Notably for our purposes, he does so throughout the remainder of the paper that begins with the necessity of identity proof on page 2 (Kripke 1971). It seems natural, then, to conclude that according to these proofs identity is *metaphysically* necessary.

That said, there is an alternative line of thought. Self-identity is a theorem of many logics. If there is a kind of necessity attached to logical validity, then it is logically necessary that everything is self-identical. Indeed, that seems to be as basic and self-evident a thought as one can think of. One might therefore take it be plausible that, insofar as we find the assumption of the necessity of self-identity compelling in the Quine-Kripke proof, we do so in the sense of *logical* necessity. And then, insofar as the proof effectively transfers the necessity of a self-identity over to an identity referred to by different means, one might think that it is logical necessity that has been transferred.

But how can this be? If logical necessity creates an opaque context, then it should be illegitimate to move from $\Box(x = x)$ to $\Box(x = y)$. Here is one suggestion: the proofs surveyed here do not conclude just with a necessitated identity, such as $\Box(x = y)$, but with a *conditional*, i.e., $(x)(y)((x = y) \supset \Box(x = y))$. This says that, for any things x, y, *if* x is identical to y, *then* this is necessarily so. That is quite different to the claim that a particular identity statement, 'Hesperus is Phosphorus', say, is necessary. This difference is underlined by a careful examination of these different claims by Burgess (2014).[29] To move from the universally quantified conditional to a claim concerning specific names requires, amongst other things, further claims about the way that proper names work (as we shall see in the next section). In short: there may be some way to make sense of the necessity implicated here as a logical necessity, if we take sufficient care to differentiate different claims in the vicinity.

3. Identity and Necessity in Metaphysics

[29] Particularly section 3.

In this section I consider two cases where the necessity of identity has played an important role in metaphysics.

3.1. *Naming and Necessity*

Kripke is famous for arguing in his *Naming and Necessity* lectures that some statements are both necessary and *a posteriori*. A central case is identity statements that are necessarily true yet not *a priori*. One metaphysically significant aspect of Kripke's conclusions is that they tease apart a metaphysical notion of necessity from an epistemic notion of *a priority*.

Here is one way to summarize the key line of argument.[30] Venus is identical to Venus. Indeed, *necessarily*, Venus is Venus. But we have several ways of referring to the planet Venus, including the proper names 'Hesperus' and 'Phosphorus'. Proper names are rigid designators; they refer to the same object in every possible world in which the object exists (and in all other worlds to nothing).[31] So, since 'Hesperus' and 'Phosphorus' each actually refer to Venus, they both refer to Venus in every world (in which Venus exists). But in every world (in which Venus exists) Venus is identical to Venus, so in every world (in which Venus exists) 'Hesperus is identical to Phosphorus' is true. So necessarily, Hesperus is Phosphorus.[32]

---

[30] See Kripke (1981, 28-29, 97-105).

[31] Kripke (1971, 146). See also Ahmed (2007, 19) for disambiguation of several candidate meanings of 'rigid designator'.

[32] This is a weak sense of necessity, according to which 'we can count statements as necessary if whenever the objects mentioned therein exist, the statement would be true' (Kripke, 1971, 137). There are important questions how to evaluate the truth

Nevertheless, names do not wear their referents on their sleeves—it is (usually) an empirical matter to discover the referent of a name. It is an empirical discovery that ‘Hesperus’ and ‘Phosphorus’ (actually) refer to the same planet. (The usual story is that the referent of ‘Hesperus’ was fixed by the description ‘the evening star’, and the referent of ‘Phosphorus’ was fixed by ‘the morning star’, although those descriptions merely fixed the referents and do not constitute the meanings of the names, long before it was discovered that the star that appears in the morning sky is the very same heavenly body, *Venus*, as appears in the evening sky.) So ‘Hesperus is Phosphorus’ is necessarily true but not *a priori*.

What is the significance of this? Kripke’s arguments for the necessary *a posteriori* present a fundamental challenge to a tendency through the history of philosophy to equate necessity with *a priority* and analyticity.[33] It is also worthwhile to set Kripke against a more recent history of debates between Quine, Barcan, and others, concerning the intelligibility of quantified modal logic.[34]

---

value of statements containing names at worlds at which the referents do not exist, but I shall not address them here. See, e.g., Fine (2005); Williamson (2002).

[33] Philosophers such as Leibniz, Locke and Hume are often cited as endorsing such an equation, with Kant, who argues for the (necessary) synthetic *a priori*, as the most notable dissenter. The equation is taken up again by logical positivists such as Ayer in the early Twentieth century. But see Kneale (1938) for an early introduction of the necessary *a posteriori*, and Leech (2019) for discussion.

[34] See in particular the transcript of the discussion of “Modalities and Intensional Languages” at the 1962 Boston Colloquium, in Barcan Marcus (1993).

To briefly sketch Quine's challenge: Start by assuming a logico-linguistic understanding of necessity, according to which necessary truths are logical and analytic truths. Since logical truths are *a priori*, so are necessary truths. It is intelligible to apply such a notion of necessity as a metalinguistic predicate. For example, just as it makes sense to apply the explicitly metalinguistic predicate 'is logically valid' to a sentence such as 'Seven is greater than five' (whether truly or not), so it makes sense to apply the metalinguistic predicate 'is necessary', i.e., '"Seven is greater than five" is necessary'. Parasitic upon this intelligible use of metalinguistic predicates, we can understand the use of a sentential operator 'necessarily', as in 'Necessarily, seven is greater than five'. However, Quine complains, if we then quantify into the scope of this operator, assuming a standard objectual reading of the quantifiers, we face trouble (Quine, 1953). If we try to say, 'Something is necessarily greater than five', i.e., '$\exists x \Box (x > 5)$', this is nonsense: it doesn't make sense to say of some thing that it is validly greater than five, and so (given the metalinguistic roots of the modal operator), it makes just as little sense to say of some thing that it is necessarily greater than five.

There are various lines of response.[35] One option is to understand the modal operator differently, by appeal to a kind of necessity that has its source in things and not in words. Hence Quine's oft-quoted warning that quantified modal logic 'leads us back into the metaphysical jungle of Aristotelian essentialism' (Quine, 1953, 174). Against this background, we can understand part of Kripke's impact. Kripke argues that identity statements are necessary but not *a priori*, and so they cannot be

[35] See, for example, Ballarin (2012); Barcan Marcus (1961); Divers (2017); Fine (1989); Smullyan (1947, 1948).

necessary in this metalinguistic, logical, sense. However we understand the necessity of identity statements—perhaps by returning to Aristotelian metaphysics—it is surely not to be in terms of analyticity or *a priority*.

We can also, once again, see the signal role played by the necessity of self-identity. That core necessity is taken for granted, and combined with considerations of how referring expressions work. As before, the issue arises due to the way that we allow ourselves to represent identities using distinct terms, such as 'Hesperus' and 'Phosphorus'. Does this detract from the *metaphysical* significance of the necessary *a posteriori*? Not obviously.[36] The point to stress is that, where the necessary *a posteriori* is introduced via identity statements, once we bracket considerations of how referring expressions work, we are left primarily with an assumption of the necessity of self-identity. *Is that a metaphysical assumption? A logical assumption? Both? Why can we take it for granted?* I would suggest that until we can answer these questions about *self-identity*, the metaphysical significance of this kind of case of the necessary *a posteriori* is left open.

3.2. Coincidence, counterparts, and contingent identity.

In *Naming and Necessity* the necessity of identity is *endorsed*. In this section, I review a case where the necessity of identity is *denied*. Even so, we will see that our choices of how to represent identity play a leading role, and that there is an important sense in which nothing is contingently the thing that it is.

---

[36] Indeed, further examples of the necessary *a posteriori,* such as the necessity of origin, do not obviously depend on representational matters, such as the function of proper names (see Kripke 1981, 114–116).

Suppose we have a statue called "Fawcett", made of a mass of bronze called "Bronze". Fawcett and Bronze share all their categorical (non-modal) properties; they have the same size and shape, they both stand on a plinth in Parliament square, and so on.[37] One way to explain why they have the same properties is that they are identical: the statue and the bronze are one and the same thing.

Nevertheless, Fawcett and Bronze appear to differ in their modal properties. If the bronze was melted down and solidified into a sphere, that would no longer be the same statue, or perhaps not a statue at all, even though it would be the very same mass of bronze.[38] So, for example, Fawcett is not possibly spherical, while Bronze *is* possibly spherical. But if Fawcett is identical to Bronze, how can that be? How can they differ in any properties, if they are the same thing? We saw earlier a law of substitutivity: $(x)(y)[(x = y) \supset (Fx \supset Fy)]$. If this includes modal predicates in its scope, as is required by the necessity of identity proof, then predicates such as "is possibly spherical" should be included too. But then, if Fawcett=Bronze, and if Bronze is possibly spherical, it follows that Fawcett is possibly spherical after all.

One response is to claim that Fawcett and Bronze are *merely contingently identical*. Since they are actually identical, they share their non-modal properties, but they might have been distinct. So, for example, there is a possible world in which Fawcett and Bronze are not identical, in which Bronze is spherical, and in which Fawcett is made from something else (e.g. a different mass of bronze, or a mass of

---

[37] I ignore temporal properties for present purposes. See Wasserman (2021) and bibliography for more on puzzles of material constitution.

[38] Is a sphere of bronze a statue? At least it seems unlikely to be a statue of Millicent Fawcett.

playdough). My case study here is to show how *counterpart theory* does an admirable job of providing a framework in which we can represent and make sense of contingent identity.[39]

Counterpart theory offers a translation of modal claims made in a language containing modal terms or operators into a first-order extensional language (Lewis, 1968). In brief, modal operators, 'possibly' and 'necessarily', are translated into quantification over worlds, e.g. 'Possibly p' becomes 'There is a world w such that p is true at w'. In the case of *de re* modal claims, possibilities concerning an individual are represented by distinct individuals at other possible worlds. For example, 'Possibly x is F' becomes 'There is a world w that contains an individual y such that y is a counterpart of x and y is F'. There is a straightforward sense here in which y is a representative of x and (some of) its possibilities.

Strictly speaking, counterpart theory provides such a translation without committing to a particular metaphysics. But it is often presented as a package along with a Lewisian metaphysics of worlds (Lewis, 1986). According to this metaphysical view, counterpart theory is not just a useful translation scheme that allows us to translate tricky intensional modal vocabulary into a more tractable extensional language, but it reflects the metaphysical reality underlying modal truths. So, for example, it is not just that we can translate *de re* talk of x being possibly F into talk of counterparts of x being F, but also that what makes it true that x is possibly F is the existence of an individual, distinct from and spatiotemporally disconnected from x,

[39] My understanding of counterpart theory has been greatly improved by discussions with Cansu Yuksel. See Yuksel (2022).

that is F, and that is similar enough to x in relevant respects to count as representing possibilities for x.

Which individuals at a world represent the *de re* possibilities for some individual x is determined by the counterpart relation. The counterpart relation is a context-sensitive similarity relation. Whether some individual y at another world is a counterpart of x depends upon which similarity relation is selected in context (see Lewis 1971). To illustrate, consider again Fawcett and Bronze. When we ask whether Fawcett could have been spherical, we are referring to Fawcett under the guise of a statue: the name 'Fawcett' is associated with the sortal *statue*. As such, the counterpart relation in this context is a "statue-similarity" relation. In this context, the counterpart(s) of Fawcett at any world are the object(s) most similar to Fawcett in the respects that are relevant for statues. So, assuming that *shape* is an important feature of statues, at a world with nothing that is remotely similar in shape to Fawcett, we would say that there is no counterpart of Fawcett. Or, in a world with a marshmallow the exact size and shape as Fawcett, with nothing else a similar shape, it would seem that there is a marshmallow counterpart of Fawcett. Likewise, when we ask whether Bronze could have been spherical, we are referring to Bronze under the guise of a mass of bronze: the name 'Bronze' is associated with the sortal *mass of bronze*. As such, the counterpart relation in this context is different. In this context, the counterpart(s) of Bronze at any world are the object(s) most similar to Bronze in the respects that are relevant for masses of chemical substances. So a sphere of bronze will count as more similar to Bronze than a statue made of marshmallow.

The upshot is that even supposing Fawcett and Bronze are *actually identical*, in some worlds they are represented by *distinct* counterparts. In some contexts, the counterpart of Fawcett at w is the most statue-similar individual to Fawcett—call it s—and the counterpart of Bronze at w is the most substance-similar individual to Bronze—call it b—such that $s \neq b$. So world w and its inhabitants represent the possibility that Fawcett and Bronze are distinct, even though they are actually identical: Fawcett and Bronze are *possibly distinct* and *contingently identical*.

This is not the place to discuss the pros and cons of counterpart theory in detail.[40] Rather, let us highlight some points of particular relevance to our discussion of the necessity of identity. First, how does counterpart theory relate to proofs of the necessity of identity? Counterpart theory translates modal language into a first-order extensional language, so we are working with a different language to those in which the necessity of identity is given a proof, i.e., languages for modal logics. If there were a proof of the necessity of identity in the language of counterpart theory, it would look very different. Lewis's 1968 counterpart theory does not have as a theorem the translation of the necessity of identity, but it is open to place different constraints on the counterpart relation to yield such a result.[41]

Counterparts and worlds are understood as *representatives* of possibility. This is made especially clear when combined with Lewisian metaphysics. In discussing *de re* possibilities for an individual, Humphrey, Lewis writes,

> How does a world, genuine or ersatz, represent, concerning Humphrey, that he exists? … [A genuine possible world] can have as a part a Humphrey of its

[40] See, for example, Fara and Williamson (2005); Mackie (2006); Wang (2015); Woodward (2012).

[41] See Varzi (2020).

> own, a flesh-and-blood counterpart of our Humphrey, a man very like Humphrey … By having such a part, a world represents, *de re*, concerning Humphrey … that he exists and does thus-and-so. (Lewis, 1986, 194)

This is an important aspect of Lewis's metaphysics: that other possible worlds and their inhabitants *represent* possibilities, and they do so by being a certain way, or by containing certain individuals with certain properties. Moreover, it is a part of a Lewisian metaphysics of counterparts that worlds do not overlap, i.e., they do not share individuals, and so the counterparts of an individual x at other worlds are not identical to x.[42] This means that the identity or distinctness at other worlds of actual individuals *never* involves those individuals themselves, but is always *represented by* individuals that are strictly speaking *distinct* from the individuals in which we are interested. If we take a step back and consider the pluriverse as a whole—the plurality of worlds and individuals in them—then there is a sense in which nothing is merely contingently identical after all. Everything is just the thing that it is and nothing else. Some individuals bear counterpart relations to other things, and in some cases those counterparts represent apparently contingent identities.

In summary, at the logico-linguistic level, counterpart theory allows us to represent true contingent identities. At the metaphysical level, counterparts are individuals that in some cases represent possibilities for individuals to be contingently identical. *But nothing is strictly speaking contingently identical*. When we consider the domain of possible individuals, everything is the thing that it is and no

[42] That's at *other* worlds. For x in world w, the counterpart of x at w is x itself (Lewis, 1968, 114).

other. Some of those things stand as representatives of other things, and in some cases represent identity and distinctness of other things, but no one individual *itself* is two or more things at another world. This is, in a way, an extraordinarily elegant way to have one's cake and eat it: one has true statements of contingent identity without really anything being merely contingently identical, since contingent identities are represented by avowedly distinct entities.[43]

Our moral is that, again, we find that the necessity of identity, or its denial, can be viewed as a combination of (a) a commitment to everything being just the thing it is and no other, and (b) important choices about how to represent identity and modality. Counterpart theory at a basic level accepts (a). It can combine (a) with a way for us to represent identity and distinctness at different possible worlds, and accommodate contingent identity, precisely by taking distinct things to represent identities.

4. Generalized Identity and Necessity

If sound, proofs of the necessity of identity show *that* identity is necessary. But *why* is identity necessary, if it is? Can we give a metaphysical *explanation* of the necessity of identity?[44] Recent work on *generalized identity* suggests so. In this section, therefore, I aim to introduce the notion of generalized identity, and to sketch how this might help to provide an explanation of the necessity of identity, and of necessity more widely.

[43] This line of thought raises issues of "advanced modalizing". See, for example, Divers (1999).

[44] On metaphysical explanation see Brenner et al. (2021).

So far in this chapter we have been concerned with *objectual* identity statements: identity-indicating expressions, such as "=" or "is", flanked by expressions standing for individuals, i.e., variables or names. But there is a large and well-recognized family of statements that look like identity statements but which do not concern (only) individuals: statements of generalized identity, or what Rayo (2013) calls "just-is" statements. For example,

1) For a thing to be a bachelor *is* for it to be an unmarried adult male.

2) For a thing to know a proposition *is* for it to truly, justifiably believe that proposition.

3) For the Atlantic Ocean to be filled with water *is* for it to be filled with H2O molecules. (Correia and Skiles, 2019, 643)

One can understand objectual identities as a special case, for example,

4) For something to be Hesperus *is* for something to be Phosphorus.

It is widely assumed that generalized identity statements are necessary, and that they correspond to necessary biconditionals. I.e., if we take a generalized identity statement to be of the form '$p \equiv q$'—read: 'for it to be the case that p is for it to be the case that q'—then if $p \equiv q$, it follows that $\Box(p \equiv q)$, and $\Box(p \leftrightarrow q)$.[45] Where

[45] '$\leftrightarrow$' here signifies the material biconditional.

generalized identity concerns predicates, $Fx \equiv_x Gx$—read: ‘for something to be F is for it to be G’—implies that $\Box(Fx \equiv_x Gx)$, and $\Box\forall x(Fx \leftrightarrow Gx)$.[46]

Can we grant the necessity of generalized identity? Why should it be so? We have seen that the necessity of objectual identity is not a foregone conclusion, but there are logical proofs available. We can employ an analogue of these proofs for generalized identity, provided we grant the necessity of self-generalized-identity, i.e. $\Box(p \equiv p)$, and a version of Leibniz’s Law:

> LL: If $p \equiv_v q$ and Φ, then $\Phi[q//p]$
>
> where $\Phi[q//p]$ results from sentence Φ by replacing one or more occurrences of p by q, with the condition that no variable that is free in $p \equiv_v q$ is bound in Φ or $\Phi[q//p]$.
>
> (Correia and Skiles, 2019, 645)

The proof then goes as follows (Leech, 2021, 901):

1.	If $p \equiv q$ and Φ, then $\Phi[q//p]$.	LL
2.	$\Box(p \equiv p)$	Necessity of self-generalized-identity
3.	$p \equiv q \supset (\Box(p \equiv p) \supset \Box(p \equiv q))$	Instance of 1 from 2
4.	$p \equiv q \supset \Box(p \equiv q)$	From 2, 3, conditional rule

The proof in itself does not provide an explanation of the necessity of generalized identity, but it is suggestive. The thought implied here (and *mutatis*

[46] See Correia and Skiles (2019, 646), Dorr (2016, 2).

mutandis for the earlier proofs) is that since p necessarily is p, and q *is* p, it must also be that q necessarily is p. I.e., $\Box(p \equiv q)$ *because* $p \equiv q$ and $\Box(p \equiv p)$.[47]

Dorr (2016) proposes that generalized identity may provide an explanation of necessity.[48]

> The claim that to be F is to be G constitutes a very satisfying *explanation* of the fact that necessarily, all and only F things are G. One might be puzzled as to why it should be necessary that everything F is G, or that everything G is F. But if to be F *is* to be G, there is nothing more to be puzzled about. (Dorr, 2016, 2)

> The fact that Hesperus is Phosphorus explains in a supremely satisfying way why it is necessary that everyone who lands on Hesperus will land on Phosphorus. Identities are excellent stopping places for explanation; they do not cry out for explanation in their own right. (Dorr, 2016, 3)

[47] One we move to an explanatory register, it is imperative that p and q in the *explanandum* be distinct: if understood as (schematic) variables that could take the same value, we would be saddled with circular explanations such as $\Box(p \equiv p)$ *because* $p \equiv p$ and $\Box(p \equiv p)$. If we want to take this explanatory route, therefore, we should specify that p and q stand for different formulas, or otherwise guard against circularity. Thank you to Kit Fine and Isaac Wilhelm for discussion.

[48] Rayo (2013) also gives an account of metaphysical modality in terms of ‘just is’-statements.

The thought here seems to be that generalized identity explains at least some necessities, i.e., the necessity of identity. But how? Our sketch of an explanation based on the proof suggests one way to go: an identity, combined with the necessity of self-identity, explains the necessity of that identity.

Dorr goes further in suggesting that we define the modal operator '$\Box$', understood as expressing metaphysical necessity, by identification with an arbitrary tautology: $\Box =_{df} \lambda p. p \equiv \top$. To be a necessity *is* to be a tautology. This, according to Dorr, provides much-needed illumination to the murky notion of metaphysical necessity.

> Philosophers have struggled to say something helpful to single out the "metaphysical" readings of modal operators from among the panoply of other readings they may bear; their efforts have not been conspicuously successful. So an explication of 'It is metaphysically necessary that φ' as 'For it to be the case that φ is for it to be the case that $\top$' would shed some welcome light on the concept of metaphysical necessity and the interest of questions formulated in terms of it. (Dorr, 2016, 41)

A self-identity such as is expressed by "$x = x$" or "$p \equiv p$" is surely tautologous if anything is. If we think of informative generalized identity statements as different ways to express a core self-identity, and if that self-identity is tautologous, then since it is plausible to take the tautology to be necessary, the generalized identity statement expressing it must also be necessary. Indeed, Dorr writes,

> For it to be necessary that everything square is square is for it to be necessary that everything square is rectangular and equilateral.
> (Dorr, 2016, 7)

This is not merely an *entailment*, but an *identification*. It is necessary that everything square is rectangular and equilateral because *that just is* for it to be necessary that everything square is square. The necessity of a self-identity is taken to be *just the same thing* as the necessity of an identity.

Rayo (2013) makes a similar move.

> Consider 'to be hot *just is* to have high mean kinetic energy' as an example. What is required of the world in order for the truth conditions of this sentence to be satisfied is that there be no difference between having high mean kinetic energy (i.e. being hot) and being hot. Equivalently: that there be no difference between being hot and being hot – a condition which is satisfied trivially.
> (Rayo, 2013, 38)

A generalized identity holds not only between *being hot* and *having high mean kinetic energy*, but also between the *identity* of being hot with having high mean kinetic energy and the *self-identity* of being hot with being hot.

These considerations from Dorr and Rayo suggest the following: $\Box(p \equiv q)$ *because* $(p \equiv q) \equiv (p \equiv p)$ and $\Box(p \equiv p)$, i.e., because the generalized identity of its being the case that p with its being the case that q *is* the generalized identity of its being the case that p with itself, and that self-generalized identity is necessary. If it is also the case that $(p \equiv q) \equiv (p \equiv p)$ *because* $(p \equiv q)$, then, if explanation is transitive, this implies the explanation sketched in line with the proof: $\Box(p \equiv q)$ *because* $p \equiv q$ and $\Box(p \equiv p)$.

Are such purported explanations of the necessity of generalized identity, in appealing to a generalized identity and a necessity, circular? No. The *explanandum* is the necessity of generalized identity: $\Box(p \equiv q)$. None of the proposed *explanantia* are the same as this, so there is no explicit circularity. The explanation appeals to a generalized identity, $(p \equiv q) \equiv (p \equiv p)$ or $p \equiv q$, but not to its *necessity*. Given the necessity of generalized identity, this generalized identity will also be necessary, and is therefore suitable to explain a necessity. But still its necessity does not play a role in the explanation.[49] What about the appeal to the necessity of self-generalized-identity? This again raises a host of familiar questions concerning the logical, metaphysical, and modal status of self-identity and tautology.[50]

References

Ahmed, Arif. 2007. *Saul Kripke*. London and New York: Continuum.

Ballarin, Roberta. 2012. “Quine on intensional entities: Modality and quantification, truth and satisfaction.” *Journal of Applied Logic* 10:238–249.

—. 2021. “Modern Origins of Modal Logic.” In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University, Fall 2021 edition.

[49] This response is similar to Hale (2002)’s response to Blackburn’s dilemma.

[50] Thank you to Julien Dutant, Nils Kürbis, and Andrew Stephenson, for feedback on the Barcan proofs. Thank you to the EuPhilo Seminar, and an audience in Hamburg, for further comments and feedback.

Barcan, Ruth. 1946a. “A functional calculus of first order based on strict implication.” *The Journal of Symbolic Logic* 11:1–16.

—. 1946b. “The deduction theorem in a functional calculus of first order based on strict implication.” *The Journal of Symbolic Logic* 11:115–118.

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

Barcan Marcus, Ruth. 1961. “Modalities and Intensional Languages.” *Synthese* 13:303–322. Reprinted in Barcan Marcus (1993), pp.5–23.

—. 1993. *Modalities: Philosophical Essays*. Oxford and New York: Oxford University Press.

Ben-Yami, Hanoch. 2018. “The Logical Contingency of Identity.” *European Journal of Analytic Philosophy* 14:5–10.

Brenner, Andrew, Maurin, Anna-Sofia, Skiles, Alexander, Stenwall, Robin, and Thompson, Naomi. 2021. “Metaphysical Explanation.” In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University, Winter 2021 edition.

Burgess, John P. 2014. “On a derivation of the necessity of identity.” *Synthese* 191:1567–1585.

Correia, F. and Skiles, A. 2019. “Grounding, Essence, and Identity.” *Philosophy and Phenomenological Research* 98:642–70.

Della Rocca, Michael. 1996. “Essentialists and Essentialism.” *Journal of Philosophy* 93:186–202.

Divers, John. 1999. “A genuine realist theory of advanced modalizing” *Mind* 108(430):217–239.

—. 2017. “De Re Modality in the Late Twentieth Century.” In M. Sinclair (ed.), *The Actual and the Possible: Modality in Modern Philosophy*, Mind Occasional Series. Oxford: Oxford University Press.

Dorr, Cian. 2016. “To be F is to be G.” *Philosophical Perspectives* 30 (Metaphysics):39–134.

Fara, Michael and Williamson, Timothy. 2005. “Counterparts and Actuality.” *Mind* 114:1–30.

Fine, Kit. 1989. “The Problem of De Re Modality.” In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, 197–272. Oxford: Oxford University Press.

—. 2005. “Necessity and Non-Existence.” In *Modality and Tense: Philosophical Papers*, 321–354. Oxford: Oxford University Press.

Fitch, Frederic B. 1949. “The problem of the morning star and the evening star.” *Philosophy of Science* 16:137–141.

Fitting, Melvin and Muskens, Reinhard. 2022. “Intensional Logic.” In Edward N. Zalta and Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. Stanford University, Winter 2022 edition.

Forrest, Peter. 2020. “The Identity of Indiscernibles.” In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University, Winter 2020 edition.

Hale, Bob. 2002. “The Source of Necessity.” *Philosophical Perspectives* 16: 299-319.

—. 2013. *Necessary Beings: An Essay on Ontology, Modality, & the Relations Between Them*. Oxford: Oxford University Press.

Kneale, Martha. 1938. “Logical and Metaphysical Necessity.” *Proceedings of the Aristotelian Society* 38(1):253-268.

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

—. 1981. *Naming and Necessity*. Oxford: Basil Blackwell. First published in G. Harman and D. Davidson (eds), *Semantics of Natural Language* (D. Reidel Publishing Co., Dordrecht, Netherlands & Boston, USA, 1972). Revised and enlarged edition first published 1980. Paperback edition 1981.

Kürbis, Nils. forthcoming. “Arthur Prior’s Proofs of the Necessity of Identity and Difference.” *History and Philosophy of Logic* 1–5.

Leech, Jessica. 2019. “Martha Kneale on Why Metaphysics Necessities Are Not A Priori.” *Journal of the American Philosophical Association* 5(4):389-409.

—. 2021. “From Essence to Necessity via Identity.” *Mind* 130:887–908.

Lewis, C. I. and Langford, C. H. 1932. *Symbolic Logic*. New York: Century Company.

Lewis, David. 1968. “Counterpart Theory and Quantified Modal Logic.” *Journal of Philosophy* 65:113–126.

—. 1971. “Counterparts of Persons and Their Bodies.” *Journal of Philosophy* 68:203–211.

—. 1986. *On the Plurality of Worlds*. Blackwell.

Mackie, Penelope. 2006. *How Things Might Have Been: Individuals, Kinds, and Essential Properties.* Oxford: Clarendon Press.

Prior, Arthur. 1962. *Formal Logic*. Oxford: Clarendon Press, 2 edition.

Quine, W. V. 1953. “Three Grades of Modal Involvement.” *Proceedings of the XIth International Congress of Philosophy* 14:65–81. Reprinted in Quine, W.V. (1996) *The Ways of Paradox*, (Random House: New York), 156–174.

Rayo, Augustín. 2013. *The Construction of Logical Space*. Oxford: Oxford University Press.

Smith, Quentin. 1995. “Marcus, Kripke, and the Origin of the New Theory of Reference.” *Synthese* 104:179–189.

Smullyan, A. F. 1947. “The Problem of Interpreting Modal Logic by W. V. Quine.” *The Journal of Symbolic Logic* 12:139–141.

—. 1948. “Modality and Description.” *The Journal of Symbolic Logic* 13:31–37.

Soames, Scott. 1995. “Revisionism about Reference: A Reply to Smith: Eastern Division Meetings of the APA Boston, December 1994.” *Synthese* 104:191–216.

Varzi, Achille C. 2020. “Counterpart theories for everyone.” *Synthese* 197(11):4691-4715.

Wang, Jennifer. 2015. “Actualist Counterpart Theory.” *Journal of Philosophy* 112:417–441.

Wasserman, Ryan. 2021. “Material Constitution.” In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University, Fall 2021 edition.

Philosophy 92:173–175.

Wiggins, David. 1965. “Identity Statements.” In R. J. Butler (ed.), *Analytic Philosophy: Second Series*, 40–71. Oxford: Basil Blackwell.

Williamson, Timothy. 2002. “Necessity Existents.” In A. O’Hear (ed.), *Logic, Thought, and Language*, 233–251. Cambridge: Cambridge University Press.

—. 2013. *Modal Logic as Metaphysics*. Oxford: Oxford University Press.

Woodward, Richard. 2012. “Counterparts.” *Philosophy Compass* 7:58–70.

Yuksel, Cansu. 2022. *Worlds, Counterparts and Modality.* PhD thesis.