

# Carnap's theory of descriptions and its problems

Jan Heylen

## Abstract

Carnap's theory of descriptions was restricted in two ways. First, the descriptive conditions had to be non-modal. Second, only primitive predicates or the identity predicate could be used to predicate something of the *descriptum*. The motivating reasons for these two restrictions that are to be found in the literature will be critically discussed. Both restrictions can be relaxed, but Carnap's theory can still be blamed for not dealing adequately with improper descriptions.

## 1 Introduction

The subject of this article is Carnap's modal logic with descriptions and its varied sorrows. Since Carnap's modal logic is not widely known and since it has some peculiar features, I will devote section 2 to an introduction to it. Also, it will be used to fix some notation. Moreover, Carnap's modal logic will be introduced against the background of more general modal logics, which can serve as background theories for arguments that are not tied to the particular modal logic devised by Carnap.

In section 3 I will introduce Carnap's theory of descriptions. It will be explained that it is faced with a problem that involves the exact formulation of the uniqueness condition. The problem, its scope, and the solutions proposed in the literature will be scrutinized and I will settle on one solution.

In section 4 I will introduce what I call 'self-predication principles', e.g. if there is one and only one queen of England, then the Queen of England is a queen. I will use self-predication to argue for the claim that Carnap's theory of descriptions inadequately copes with improper descriptions. Also, I will set the stage for a discussion of two arguments by Føllesdal for the claim that modal distinctions collapse in Carnap's modal logic with descriptions. This will be done by showing that certain self-predication principles are valid, whereas others are not.

In section 5 I will discuss Føllesdal's first collapse argument, which uses an instantiation of a restricted self-predication principle. The problem with this argument is that it assumes the principle of the necessity of identity, which is invalid in a Carnapian interpretation. I will also discuss the relations between Føllesdal's argument and a collapse argument and a slingshot argument by Quine. It will turn out that the connections between collapse arguments and slingshot arguments is tighter than has been recognized by some.

In section 6 I will discuss Føllesdal’s second collapse argument, which uses an instantiation of an unrestricted self-predication principle. I will then formulate another collapse argument that is in some interesting respects weaker than Føllesdal’s (or Quine’s, for that matter). The problem with these arguments is that the self-predication principle is invalid. Moreover, they violate a certain restriction by Carnap. This restriction and its alleged motivation will be discussed as well.

## 2 Carnapian modal logic

In this section I will introduce Carnapian intensional logic against the background of two other intensional logics. The first logic is a very general logic, viz. a ‘contingent identity system’, known as **S+CI** (‘Contingent Identity’). On the model-theoretic level, its main feature is that the variables range over ‘allowable’ individual concepts. The second logic is an extension of the first logic that will be called **S5+IC** (‘Intensional Concepts’). On the model-theoretic level, its main feature is that the variables range over *all* individual concepts. The third logic is a Carnapian intensional logic, **C**. On the model-theoretic level, its main features are that variables range over *all* individual concepts and that the class of worlds includes (in a sense to be specified) *all* possible worlds. It should be clear that a **C**-model is a **S+CI**-model, but not vice versa. By moving from a textbook system **S+CI** to a not so well-known system **C**, one can gradually get to know the characteristic features of the latter special logic.

Let  $\mathcal{L}$  be a first-order language with identity and with the modal operator  $\Box$ . Well-formed terms, formulas and sentences are defined as usual. The **S+CI**-interpretation of  $\mathcal{L}$  can be found in a textbook [7, p. 333-334]. An **S+CI**-model for  $\mathcal{L}$  is a quintuple  $\langle W, R, D, I, V \rangle$  with  $W$  a non-empty set (of ‘possible worlds’),  $R$  a two-place (‘accessibility’) relation on  $W$ ,  $D$  is a non-empty set (the ‘domain’),  $I$  is a set of functions from  $W$  to  $D$  (the ‘allowable individual concepts’) and  $V$  is a function defined as follows: if  $P^n$  is a  $n$ -place predicate, then  $V(P^n, w) \subseteq D^n$ . The assignment function  $a$  is a function from the set of variables to the set of the allowable individual concepts.

Given an **S+CI**-model  $\mathcal{M}$ , an assignment function  $a$  and a well-formed term  $t$  of  $\mathcal{L}$ , let  $V_{\mathcal{M}}(t, w, a)$  be a function defined as follows: if  $x$  is a variable and if  $a(x) = i$ , then  $V_{\mathcal{M}}(x, w, a) = a(x)(w) = i(w)$ . Given an **S+CI**-model  $\mathcal{M}$ , an assignment function  $a$  and  $\phi$  a well-formed formula of  $\mathcal{L}$ , let  $V_{\mathcal{M}}(\phi, w, a)$  be a function defined as follows: if  $P^n$  is a  $n$ -place predicate of  $\mathcal{L}$  and if  $t_1, \dots, t_n$  are well-formed terms  $\mathcal{L}$ , then  $V_{\mathcal{M}}(P^n(t_1, \dots, t_n), w, a) = T$  iff  $\langle V_{\mathcal{M}}(t_1, w, a), \dots, V_{\mathcal{M}}(t_n, w, a) \rangle \in V(P, w)$  and  $V_{\mathcal{M}}(t_1 = t_2, w, a) = T$  iff  $V_{\mathcal{M}}(t_1, w, a) = V_{\mathcal{M}}(t_2, w, a)$ ; the clauses for  $\neg$  and  $\wedge$  are as can be expected;  $V_{\mathcal{M}}(\forall x\phi, w, a) = T$  iff  $V_{\mathcal{M}}(\phi, w, a[i/x]) = T$  for every  $i \in I$ ;  $V_{\mathcal{M}}(\Box\phi, w, a) = T$  iff  $V_{\mathcal{M}}(\phi, w', a) = T$  for every  $w' \in W$  such that  $wRw'$ . Validity and model-theoretic consequence are defined as usual.

**S+CI** consists of the following axioms, axiom schemes and rules (with  $\phi, \psi \in \mathcal{L}$ , unless otherwise stipulated) [7, p. 333-334]:

**S'** If  $\mathbf{S} \vdash \phi$  with  $\phi$  belonging to the language of modal propositional logic, and if  $\psi_1, \dots, \psi_n \in \mathcal{L}$ , then  $\mathbf{S} + \mathbf{CI} \vdash \phi[\psi_1/p_1, \dots, \psi_n/p_n]$

**BF**  $\forall x \Box \phi \rightarrow \Box \forall x \phi$

**SI**  $x = x$

**CI**  $x = y \rightarrow (\phi \rightarrow \phi(y/x))$  for all  $\phi \in \mathcal{L} \setminus \{\Box\}$

**UI**  $\forall x \phi(x) \rightarrow \phi(y/x)$

**N** If  $\mathbf{S} + \mathbf{CI} \vdash \phi$ , then  $\mathbf{S} + \mathbf{CI} \vdash \Box \phi$

**MP** If  $\mathbf{S} + \mathbf{CI} \vdash \phi$ , and  $\mathbf{S} + \mathbf{CI} \vdash \phi \rightarrow \psi$ , then  $\mathbf{S} + \mathbf{CI} \vdash \psi$

**UG** If  $\mathbf{S} + \mathbf{CI} \vdash \psi \rightarrow \theta$ , then  $\mathbf{S} + \mathbf{CI} \vdash \psi \rightarrow \forall x \theta$  if  $x$  is not free in  $\psi$

The above theory is sound and complete for the given interpretation. Note that, if Contingent Identity systems had contained the unrestricted principle of the substitutivity of identicals, then **NI** would have been a theorem of them [7, 313].

**NI**  $\forall x \forall y (x = y \rightarrow \Box x = y)$

As one can readily verify, **NI** is invalid under the interpretation described above, so it is a good thing that the theory does not contain the unrestricted principle of the substitutivity of identicals.

Now that we are familiar with **S+CI** we can start to familiarize ourselves with **S5+IC**. The language stays the same. The interpretation differs however ([7, p. 331-332, 335]; [11, p. 632-664]). A **S5+IC**-model is a **S+CI**-model that satisfies two requirements: first, the accessibility relation is an equivalence relation and, second, the class of allowable individual concepts is the class of *all* functions from the set of possible worlds to the domain. The definition of the denotation of a well-formed term at a given world under a given assignment to the variables and the definition of the satisfaction of a well-formed formula at a given world under a given assignment to the variables remain the same. **S5+IC** is exactly like **S5+CI**, except for the facts that **BF** is a theorem and not an axiom of quantified **S5** [7, p. 246-247] and that it contains the following axiom schemes:

**rGF**  $\Box \exists x \phi \rightarrow \exists x \Box \phi$  for all  $\phi \in \mathcal{L} \setminus \{\Box\}$

**CO**  $\forall x \exists y (x = y \wedge \forall z (x = z \rightarrow \Box (x = y \rightarrow x = z)))$

The above axiomatization is due to Kripke [12].<sup>1</sup> Given that every **S5+IC**-model is also a **S+CI**-model, the soundness proof of the other principles of

<sup>1</sup>See [7, p. 332]. The proof only works if the well-formed formula within the scope of the existential quantifier and the modal operator is itself purely extensional. Unfortunately, it is claimed in [7, p. 332, p. 348, n.6] that the unrestricted Ghilardi Formula is valid too. Shapiro pointed the invalidity of the unrestricted Ghilardi Formula out to Martí [13, n17]. His example made use of individual constants, which are not included in the language under consideration, but it easy to adjust his example so as to work for  $\mathcal{L}$ .

**S5+IC** is textbook material, so one can conclude that **S5+IC** is sound for the given interpretation. Kripke [12, p. 72] claims that the proof of the completeness theorem is reducible to the one for **S+IC**. Interestingly, if the class of allowable individual concepts is the class of all individual concepts, then one cannot prove completeness if the theory is as strong as or weaker than quantified **S4.3.1** [7, p. 336-342] or quantified **B**.

It was already claimed that one can prove a stronger but still restricted substitutivity of identicals principle in **S+CI**. One can prove in **S5+IC** a weak unrestricted substitutivity of identicals principle, namely the principle of the substitutivity of necessary identicals:

$$\mathbf{SNI} \quad \Box x = y \rightarrow (\phi(x) \rightarrow \phi(y/x))$$

The above principle was endorsed by Carnap [3, p. 177].

After having described **S+CI** and **S5+IC**, it is easier to present Carnap's system **C**.<sup>2</sup> The **C**-model is a maximal **S5+IC**-model, i.e. a **S5+IC**-model that is such that  $|D| = |\mathbb{N}|$ ;  $R = W^2$ ; for every predicate  $P^n$  of  $\mathcal{L}$  (other than identity) and for every  $D^* \subseteq D^n$ , there is a world  $w \in W$  such that  $V(P^n, w) = D^*$ ; for all  $w, w' \in W$ , if  $V(P^n, w) = V(P^n, w')$  for every  $P^n$  of  $\mathcal{L}$ , then  $w = w'$ .

Just as **S+CI**-models have *a* class of individual concepts and **S5+IC**-models have *the* class of *all* individual concepts, so **S5+IC**-models have *a* class of possible states of the world and Carnap's original model has *the* class of *all* possible worlds [3, p. 184].<sup>3</sup>

The maximality of the **C**-model has interesting consequences for soundness and completeness. It can be proved that, for any proposition letter and the **C**-model  $\mathcal{M}$  (restricted to the language of propositional modal logic) and world  $w$ , it is the case that  $V_{\mathcal{M}}(\Diamond p, w) = T$ . As a result, the **S'** rule is not sound anymore, since it would follow from the hypothetical soundness **S'** and the noted fact about **C** that

$$V_{\mathcal{M}}(\Diamond (P^n(x_1, \dots, x_n) \wedge \neg P^n(x_1, \dots, x_n)), w, a) = T$$

for any possible world  $w$  and assignment  $a$ .

<sup>2</sup>There are three works by Carnap in which he discusses model-theory for modal logic: [2], [3], and [4]. Readers who are interested in the exact relations between those three theories should pay close attention to [3, p. 183, n. 3] and [3, p. 891, n.10; p. 892, n. 11]. It will become clear that the systems discussed in those three works differ greatly. The system discussed in this section is close to the one presented in his latest work. The main differences are that the language of the system in this section is first-order and not type-theoretical and that it does not contain individual constants, functions, descriptions and lambda terms. Moreover, in order to give a more modern, Kripke-style model theory, I will borrow from [15, p. 51-52]. See [18] for more on the peculiar properties of Carnap's modal logic. In [18] Carnap's propositional modal logic is referred to as **C**, whereas in this paper **C** is used to refer to his first-order modal logic.

<sup>3</sup>Carnap intended his modal system to be a system of logical modality. That is why he designed the range of the box operator as what is in a sense the class of all first-order models for the given language. Whether he has succeeded or not, depends on how one answer certain vexed questions, such as: 'Should it be a logical truth that there exists more than or even infinitely many individuals?'. Questions about the intended interpretation will be set aside in this paper, because there are other questions to be answered as well.

$\mathbf{C}$  is not completely and effectively axiomatizable. Suppose that it were, and suppose that  $\mathbf{C} \not\vdash \phi$  for some closed  $\phi \in \mathcal{L} \setminus \{=, \Box\}$ . Ergo,  $\phi$  is not a theorem of predicate logic without identity. By completeness, there must be a model of predicate logic without identity that makes  $\phi$  false and, hence,  $\neg\phi$  true. By the Löwenheim-Skolem theorem, there is a model with an enumerably infinite domain that makes  $\neg\phi$  true. One can turn that model into a world in the maximal  $\mathbf{C}$ -model  $\mathcal{M}$ . It follows that for every assignment  $a$ , world  $w$ ,  $V_{\mathcal{M}}(\Diamond\neg\phi, w, a) = T$ . Consequently,  $\mathbf{C} \vdash \Diamond\neg\phi$ . Since  $\mathbf{C}$  is assumed to be effectively axiomatizable, its theorems are recursively enumerable. A subset consists of the theorems of the form  $\Diamond\neg\phi$  for  $\mathbf{C} \not\vdash \phi$  and  $\phi \in \mathcal{L} \setminus \{=, \Box\}$ . But then one also has a recursive enumeration of the non-theorems of non-modal first-order logic without identity, which is in contradiction with the undecidability of the latter [1, p. 271].

In this section I have described three different logics:  $\mathbf{S+CI}$ ,  $\mathbf{S5+IC}$ , and  $\mathbf{C}$ . In the first logic the quantifiers range over *a* class of possible worlds and the modal operators range over *a* class of individual concepts. In the second logic the quantifiers range over *the* class of all individual concepts, whereas the modal operators still range over *a* class of possible worlds. In the third logic the quantifiers also range over *the* class of all individual concepts, but now the modal operators range over *the* class of *all* possible worlds. The first two logics are axiomatizable, whereas the third logic is not.

All three logics were formulated in a language that did not contain any individual constants or description operators. In the next section we will see what happens if one adds individual constants and description operators to the language.

### 3 Carnapian theories of descriptions

Carnap did not only devise a modal first-order logic, he also came up with a theory of descriptions [3, p. 37]. Let  $\mathcal{L}^*$  be identical to the language of modal first-order logic,  $\mathcal{L}$ , except for the fact that it contains a denumerably infinite set  $\mathcal{C}$  of individual constants  $c_n$  for all  $n \in \mathbb{N}$ , and a term-forming operator  $\iota$ . If  $\phi$  is a well-formed formula, then  $\iota x\phi$  is a well-formed term.<sup>4</sup> The  $\mathbf{S+CI}$ -models, the  $\mathbf{S5+IC}$ -models, and the  $\mathbf{C}$ -model are as before, except for the facts that (i) they now contain a element  $d^* \in D$ , a distinguished element of the domain, and (ii) the denotatum of an individual constant is an element of the domain ( $V(c) \in D$ ), and the denotatum of the distinguished individual constant is the distinguished element of the domain ( $V(c_0) = d^*$ ), and (iii) they have a denotation clause for descriptions.

Before stating Carnap's theory of descriptions, it is useful to give an abbreviatory definition:  $1\phi(x) =_{df} \forall y (\phi(y) \leftrightarrow y = x)$ . With the help of this definition one can formulate Carnap's theory of descriptions as follows:<sup>5</sup>

<sup>4</sup>Note that the well-formed terms and formulas of  $\mathcal{L}^*$  are to be given a simultaneous recursive definition instead of successive recursive definitions.

<sup>5</sup>Carnap was indifferent to the the choice between either giving a contextual and eliminative

$$\mathbf{C}\iota \psi (\iota x (x)) \leftrightarrow (\exists x (1\phi (x) \wedge \psi (x)) \vee (\neg \exists x 1\phi (x) \wedge \psi (c_0)))$$

Carnap infamously restricted the range of possible instantiations of the above axiom scheme. He wrote the following:

In order to avoid certain complications, which cannot be explained here, it seems advisable to admit in  $[\mathbf{C}]$  only descriptions which do not contain  $[\Box]$ . But any description may, of course, occur within the scope of  $[\Box]$ . The smallest matrix in which a description occurs (in the primitive notation) is always a nonmodal context, because the description must be an argument expression either of a primitive predicator constant or of  $[=]$ . This smallest matrix is then taken as the context of ‘ $-\iota x (\dots x \dots) -$ ’, which can be eliminated into [the right-hand side  $\mathbf{C}\iota$ ]

So a more accurate statement of Carnap’s description axiom is the following:

$$\mathbf{C}\iota \psi (\iota x (x)) \leftrightarrow (\exists x (1\phi (x) \wedge \psi (x)) \vee (\neg \exists x 1\phi (x) \wedge \psi (c_0)))$$

for all  $\phi \in \mathcal{L}^* \setminus \{\Box\}$  and for all atomic  $\psi \in \mathcal{L}^*$

In this section only the first restriction, viz. the restriction on the descriptive conditions, will be discussed. The second restriction is a topic for a later section.

The reason Carnap gives for this restriction remains obscure: he only speaks about ‘certain complications’. As a result, there was some speculation about what these ‘complications’ could be. Føllesdal gave the following reason:

As a matter of fact,  $[\phi \leftrightarrow \Box\phi]$  is not derivable in Carnap’s system. However, what prevents it from being derivable is a restriction which Carnap puts on definite descriptions in his system, to the effect that no description may contain a modal operator. No justification is given for that restriction; all Carnap says is that “in order to avoid certain complications, which cannot be explained here, it seems advisable to admit in  $[\mathbf{C}]$  only descriptions which do not contain  $[\Box]$ ”. If one lifts the restriction one sees what these complications are: modal distinctions collapse. [5, p. 150]

Discussion of Føllesdal’s reason will be postponed until the sixth section. Martí believes she knows she knows Carnap’s motivation [13, p. 583-585]. She puts forward the following explanation:

The bottom line is that if there is more than one individual in the universe of discourse the uniqueness condition cannot be satisfied, and thus all descriptions with inner modalities [are co-referential with  $c_0$ ]. I think that this is the problem Carnap is trying to avoid when he talks about the “complications which cannot be explained”. It is a technical problem and it is given an *ad hoc* solution. [13, 584-585]

---

definition of well-formed formulas containing the description operator  $\iota$  or giving an axiom scheme for  $\iota$  and adding  $\iota$  to the language of first-order modal logic [3, p. 37]. But one could also have eliminated individual constants [7, p. 184]. Moreover, in this article the possibility of languages with uneliminable descriptions will also be considered, so the second option is the preferred one.

Her argument goes as follows. Let  $\phi$  be a well-formed formula belonging to  $\mathcal{L}^* \setminus \{=, \Box\}$  (Martí spoke of ‘properties’.) What is to be proved, is that for every **C**-model  $\mathcal{M}$  with  $|D| \geq 2$  (and all **C**-models have a domain with the same cardinality as the set of natural numbers), for every world  $w$ , and for every assignment  $a$ ,

$$V_{\mathcal{M}}(\exists x \forall y (\Box \phi(y) \leftrightarrow y = x), w, a) = F.$$

The proof is by reductio ad absurdum. Suppose that there is a **C**-model  $\mathcal{M}$  with  $|D| \geq 2$ , a world  $w$ , and an assignment  $a$ , such that

$$V_{\mathcal{M}}(\exists x \forall y (\Box \phi(y) \leftrightarrow y = x), w, a) = T.$$

It follows that there is an  $i$  such that

$$V_{\mathcal{M}}(\forall y (\Box \phi(y) \leftrightarrow y = x), w, a[i/x]) = T.$$

Let  $i(w) = d$ . Now either  $\phi$  is a theorem of non-modal first-order logic without identity, or it is not. In the first case, it will be satisfied by any element of the domain in every world. Consequently, there will be an individual concept  $i'$  and an element  $d'$  such that  $d' \neq d$ ,  $i'(w) = d'$ , and  $V_{\mathcal{M}}(\Box \phi(y), w, a[i/x; i'/y]) = T$ , but  $V_{\mathcal{M}}(x = y, w, a[i/x; i'/y]) = F$ . Since the variables range over *all* individual concepts, this suffices to contradict the assumption. Suppose  $\phi$  is not a theorem of the relevant kind. Then there will be a world  $w' \in W$ , and individual concept  $i'$ , and an element  $d'$  such that  $i'(w) = d$ ,  $i'(w') = d'$  and  $V_{\mathcal{M}}(\phi(y), w', a[i/x; i'/y]) = F$ . The latter follows from the maximality of every **C**-model. Since the variables range over *all* individual concepts, this again suffices to contradict the assumption. This establishes the conclusion.

The scope of Martí’s result is restricted in two ways. First, there are more possible descriptive conditions in which a box operator occurs than the ones of the form  $\Box \phi$ . Hence, Martí’s argument supports only a more limited restriction than Carnap’s. Take for instance the descriptive condition  $F(y) \wedge \neg \Box F(y)$ . Now consider a **C**-model that is such that there is an element  $d \in D$  such that there is a world  $w$  such that  $V(F, w) = d$ . There is a class  $I' = \{i' \in I' \mid i'(w) = d\}$ . It is clear that all and only  $i'$  satisfy  $F(x)$  at  $w$  and, since all individual concepts satisfy  $\neg \Box F(x)$ , it follows that

$$V_{\mathcal{M}}(\exists x \forall y ((F(y) \wedge \neg \Box F(y)) \leftrightarrow y = x), w_2, a) = T.$$

At this point one might wonder whether Martí’s reason for Carnap’s restriction is correct, or one might conclude that Carnap could have introduced a weaker restriction. Second, Martí’s argument cannot be repeated for **S5+IC**-models and *ipso facto* not for **S+CI**-models, since the proof depends on the maximality property of **C**-models. There is a **S5+IC**-model with at least two elements in its domain and a descriptive condition  $\Box \phi(y)$  such that the uniqueness condition as formulated by Carnap *is* fulfilled. Let  $D = \{d^*, d_1\}$ ,  $W = \{w_1, w_2\}$ ,  $V(F, w_1) = d_1$ ,  $V(F, w_2) = d^*$ ,  $V(G, w_1) = \emptyset$ ,  $V(G, w_2) = d_1$ . Let  $\phi$  be  $(F(y) \vee G(y))$ . It can be verified that

$$V_{\mathcal{M}}(\exists x \forall y (\Box \phi(y) \leftrightarrow y = x), w_1, a) = T.$$

Martí himself tried to fix the problem by replacing  $1\phi(x)$  in  $\mathbf{C}\iota$  by  $i\phi(x) =_{df} \forall y (\Box P(x) \leftrightarrow \Box x = y)$ , when  $\phi$  is of a problematic form. As a result, the descriptive condition is required to be satisfied by a unique individual concept rather than an unique individual ([13, p. 585], [11, at p. 630]). Moreover, Martí had to retain  $\mathbf{C}\iota$  for descriptions with non-modal descriptive conditions, since it is clearly to though-minded to expect intensional uniqueness for extensional descriptive conditions.

The solution proposed by Martí does not work however. Suppose for a reductio ad absurdum that there is a  $\mathbf{C}$ -model, a world  $w$  and an assignment  $a$  such that  $V_{\mathcal{M}}(\exists x \forall y (\Box \phi(y) \leftrightarrow y = x), w, a) = T$ .  $\phi$  is subject to the same restrictions as above. As Martí noted, this implies that there is one and only one individual concept  $i$  satisfying  $\Box \phi(x)$ . As a consequence, there is no  $w \in W$  such that  $\phi(x)$  is true of  $d$  and  $d'$  (with  $d \neq d'$ ) at  $w$ , since otherwise there would be two individual concepts  $i, i'$  satisfying  $\Box \phi(x)$ , because one can define  $i'$  as follows:  $i'(w') = i(w')$  for all  $w' \in W$ , except that  $i'(w) = d'$ , whereas  $i(w) = d$ . But consider the world  $w'$  such that all predicate symbols in  $\phi(x)$  are assigned the empty set at  $w'$ . This world is in  $W$ , since the model is maximal. It can be proved that at  $w'$   $\phi(x)$  is either satisfied by all assignments  $a$  or by no  $a$ . If the first, then there cannot be a unique individual concepts satisfying  $\Box \phi(x)$ . If the second, there is no individual concept at all satisfying  $\Box \phi(x)$ . Either way, the reduction assumption is false. The proof is by induction on the complexity of  $\phi$ . The atomic case is trivial:  $\phi$  is by the choice of  $w'$  satisfied by no assignment. I leave the other cases to the reader.

Kremer noted that (i) one can have extensional uniqueness for descriptive conditions as well and, as a result, one only needs one description axiom (or contextual definition), and (ii) one can (still) avoid Martí's result. His proposal is to replace  $1\phi(x)$  in  $\mathbf{C}\iota$  by  $!\phi(x) =_{df} \phi(x) \wedge \forall y (\phi(y) \rightarrow y = x)$ . In classical first-order logic the definition of  $!\phi(x)$  and the definition of  $1\phi(x)$  are provably equivalent, but not so in Carnapian quantified modal logic, since the proof involves the unrestricted classical substitution principle. The resulting description principle is:

$$\mathbf{K}\iota \quad \psi(\iota x(x)) \leftrightarrow (\exists x (!\phi(x) \wedge \psi(x)) \vee (\neg \exists x !\phi(x) \wedge \psi(c_0))) \text{ for all } \phi \in \mathcal{L}^* \text{ and} \\ \text{for all atomic } \psi \in \mathcal{L}^*$$

Now there are no more restrictions on the descriptive conditions. Martí's argument depended crucially on there being an equivalence in the uniqueness conditions, but now it has been replaced by a left-to-right arrow, the argument breaks down. Moreover, one can prove that there is a  $\phi$ , a  $\mathbf{C}$ -model, a world  $w$  and an assignment  $a$  such that  $V_{\mathcal{M}}(\exists x \forall y (\Box \phi(y) \rightarrow y = x), w, a) = T$ . Let  $\phi$  be  $F(x) \vee \neg \exists z F(z)$ . Let  $w$  be such that  $V(F, w) = d$ . Let  $a(i x)$  and  $i$  is defined as follows:  $i(w) = d$  and for all  $w' \in W$  such that  $w' \neq w$ , if  $V(F, w') \neq \emptyset$ , then  $i(w') \in V(F, w')$ . This example is not due to Kremer. It could not have been, since he did not discuss  $\mathbf{C}$ -models but rather  $\mathbf{S5+IC}$ -models. Anyway, since it works,  $\mathbf{K}\iota$  will be used from now on. As a result, the denotation clause of description terms reads:  $V_{\mathcal{M}}(\iota x \phi, w, a) = d$  if for some



$i (\in I)$ ,  $V_{\mathcal{M}}(\phi, w, a[i/x]) = T$  and for every  $i$ , if  $V_{\mathcal{M}}(\phi, w, a[i/x]) = T$ , then  $i(w) = d$ ;  $V_{\mathcal{M}}(\iota x\phi, w, a) = d^*$  otherwise.

In this section Carnap's theory of descriptions was presented. Carnap restricted the allowable descriptive conditions to non-modal conditions. Føllesdal speculated that the reason for this restriction was that one could otherwise derive a modal collapse. This will be the topic for the sixth section. Martí pointed out that there was a problem with the uniqueness condition as it was formulated by Carnap: the uniqueness condition was never satisfied. This problem is limited in two ways. First, the problem identified by Martí justifies strictly speaking only a more relaxed restriction than Carnap's. Second, the problem is limited to **C**. Martí proposed an alternative, but it was proved that it does not solve the problem. Kremer proposed yet another alternative, and this alternative does the job.

Before closing this section, I would like to point out a fact that will prove useful in discussing Føllesdal's argument. In the previous section universal instantiation was restricted to variables. One could also consider a generalized universal instantiation axiom scheme:

**GUI**  $\forall x\phi \rightarrow \phi(t/x)$  for any  $t \in \mathcal{L}^*$

The above axiom scheme is sound under the last two interpretations described in the previous section. The reason is that, if  $\phi$  is true of all individual concepts, then it is also true of the individual concept determined by an individual concept, namely a constant function from the domain to the possible worlds, and the individual concept determined by a description term, namely the function that assigns to all worlds the unique object that satisfies the descriptive condition at that world if there is such an object at that world and  $d^*$  to all other worlds. There is no guarantee that the individual concept that is determined by an individual constant or a description is included in the set of allowable individual concepts, so **GUI** is not sound under the first interpretation.

Since two additional principles, **K $\iota$**  and **GUI**, have been introduced, it might be a good idea to introduce some abbreviations. Let **S+CI $\iota$**  abbreviate **S+CI+K $\iota$** , let **S+ICt** abbreviate **S+CI+GUI+K $\iota$** , and let **Ct** abbreviate **C+GUI+K $\iota$** .

## 4 Self-predication principles

This section will be about a class of principles that I dub 'self-predication principles', i.e. principles that stipulate that, if certain conditions are fulfilled, then the (possibly complex) predicate that forms the descriptive condition of a description term can be combined with the description term formed by that predicate so as to express a true proposition. These principles will be discussed for two reasons. First, as I will argue in this section, they can be used to level an argument against Carnap's theory of descriptions (independently of how one formulates the uniqueness condition). Second, they may teach us something

about Carnap’s restrictions. Third, they may illuminate Føllesdal argument for Carnap’s first restriction.

Carnap’s theory of descriptions is inadequate in its treatment of sentences that involve improper description terms that are self-predicated. The following sentences express facts:

- (a) There is no king of France.
- (b) There is no president of England
- (c) Everything is either bald or hairy, but not both.<sup>6</sup>

The following claim seems to express an intuitive truth:

- (d) Either one accepts both (i) and (ii), or one rejects both (i) and (ii).
  - (i) The bald king of France is bald.
  - (ii) The hairy president of England is hairy.

Since (a) and (b) are true, ‘the bald king of France’ and ‘the hairy president of England’ refer to  $d^*$ . Moreover, since (c) is true,  $d^*$  is either bald or hairy, but not both. If one accepts both (i) and (ii), then one deems (i) and (ii) to be true, and hence one is committed to ascribing ‘bald’ and ‘hairy’ to  $d^*$ . If one rejects both (i) and (ii), then one deems (i) and (ii) to be false, and hence one is again committed to ascribing ‘non-bald’ and ‘non-hairy’ to  $d^*$ . The conclusion is that someone who accepts Carnap’s theory of descriptions (independently of how one formulates the uniqueness condition), has to reject (d) and, hence, assign different truth-values to (i) and (ii). But (d) seems eminently plausible. To conclude, Carnap’s theory of improper descriptions is fatally flawed.

Turning from Carnap’s theory of descriptions to his restrictions, it is worth recalling that there are two of them: (i) no modal operators are allowed in a descriptive condition; (ii) no predicates other than primitive predicates or the identity predicate can be used to predicate something of the *descriptum*. Obviously, in the case of self-predication a restriction on the descriptive conditions of description terms is at the same time a restriction on the predicates that can be used to predicate something of the *descriptum*, and vice versa. When discussing the instrumental value of self-predication principles for understanding Carnap’s restrictions on his description principle (or Kremer’s), it is convenient to have a look at self-predication principles that respect the restrictions and ones that don’t:

**AtomSelfPred**  $\exists x!\phi(x) \rightarrow \phi(\iota x\phi(x))$  for all atomic  $\phi \in \mathcal{L}^*$

**NonModSelfPred**  $\exists x!\phi(x) \rightarrow \phi(\iota x\phi(x))$  for all  $\phi \in \mathcal{L}^* \setminus \{\square\}$

**S-UnrestrSelfPred**  $\exists x!\phi(x) \rightarrow \phi(\iota x\phi(x))$  for all  $\phi \in \mathcal{L}^*$

---

<sup>6</sup>I will ignore vagueness. It is not essential for the argument to work. Also, I will allow some abuse of common language, since it is not really appropriate to call a knife bald. Again, this is not essential for the argument to work.

**W-UnrestrSelfPred**  $\vdash \exists x! \phi(x) \Rightarrow \vdash \phi(\iota x \phi(x))$  for all  $\phi \in \mathcal{L}^*$

The first two principles are restricted self-predication principles, whereas the last two principles are unrestricted self-predication principles.

Both **AtomSelfPred**, which obeys the first of Carnap's restrictions, and **NonModSelfPred**, which obeys the second (and, ipso facto, also the first) of Carnap's restrictions, are **K+CI**-valid. The proof of **AtomSelfPred** is easy, since if there is at least one individual concept  $i$  with  $i(w) = d$  such that  $V_{\mathcal{M}}(!P^n(\dots, x, \dots), w, a[i/x]) = T$ , then  $V(\iota x P^n(\dots, x, \dots), w, a) = d$  and  $\langle \dots, V(\iota x P^n(\dots, x, \dots), w, a), \dots \rangle \in V(P^n, \cdot)$ . I will only sketch the proof of **NonModSelfPred**. Use the validity of **CI** to establish that, if

$$V_{\mathcal{M}}(\phi(x) \wedge \forall y(\phi(y) \rightarrow y = x), w, a[i/x]) = T$$

for some  $i$ , then

$$V_{\mathcal{M}}(\forall y(y = x \rightarrow \phi(y)), w, a[i/x]) = T.$$

By stipulation,  $V_{\mathcal{M}}(\iota x \phi(x), w, a) = i(w)$ . Ergo,  $V_{\mathcal{M}}(\phi(\iota x \phi(x)), w, a) = T$ .

It should be clear that both unrestricted self-predication principles would also have been provable if it were not for the second restriction imposed by Carnap. Indeed, the antecedents of the each of those principles imply the disjunctions of those antecedents and  $\neg \exists x! \phi(x) \wedge \phi(c_0)$ , whence by the unrestricted Carnap-Kremer description principle it follows that  $\psi(\iota x \phi(x))$ . Fortunately, Carnap did impose his second restriction, for **S-UnrestrSelfPred** and **W-UnrestrSelfPred** are respectively **Ct**-invalid and **Ct**-unsound. To prove this, I will prove that **W-UnrestrSelfPred** is **Ct**-unsound, since **W-UnrestrSelfPred** is the weakest of the both unrestricted self-predication principles, because it follows from **S-UnrestrSelfPred** by **MP**, whereas **S-UnrestrSelfPred** does not follow from **W-UnrestrSelfPred**.

**Theorem 4.1.** *W-UnrestrSelfPred is Ct-unsound.*

*Proof.* The proof breaks down into two parts. First, it needs to be established that for a certain well-formed formula  $\phi$ ,  $\models_{\mathbf{C}} \exists x! \phi(x)$ . Second, it needs to be proved that there is a **Ct**-model such that  $\phi(\iota x \phi(x))$  is not true at a certain world in that **Ct**-model. As for the first part, note that  $(F(c_3) \wedge \Box c_1 = c_1) \vee (\neg F(c_3) \wedge \Box c_2 = c_2)$  is valid. Now either  $V_{\mathcal{M}}(F(c_3), w) = T$ , or it is not. Let  $i(w') = V(c_1) = d$  for all  $w' \in W$ , and  $i'(w') = V(c_2) = d'$  for all  $w' \in W$ . Then in the first case

$$V_{\mathcal{M}}(F(c_3) \wedge \Box x = c_1, w, a[i/x]) = T,$$

and in the second case,

$$V_{\mathcal{M}}(\neg F(c_3) \wedge \Box x = c_2, w, a[i'/x]) = T.$$

Either way, there is an  $i''$  such that

$$V_{\mathcal{M}}(((F(c_3) \wedge \Box x = c_1) \vee (\neg F(c_3) \wedge \Box x = c_2)), w, a[i''/x]) = T.$$

Let us abbreviate the complex formula as  $F^\dagger(x)$ . Moreover, for every  $i'''$  for which  $V_{\mathcal{M}}(F^\dagger(x), w, a[i'''/x]) = T$ , it is the case that  $i'''(w) = i''(w)$ . Again, either  $V_{\mathcal{M}}(F(c_3)) = T$ , or it is not. In the first case  $i''(w) = d$  and so must  $i'''$ , because if  $i'''(w) \neq d$ , then  $V_{\mathcal{M}}(\Box x = c_1, w, a[i'''/x]) = F$ . The second case is analogous. Nothing hinges on the choice of  $w$ . This concludes the first part of the theorem.

Consider a **C**-model  $\mathcal{M}$  that is such that  $d_3 \in V(F, w_1)$ , it is not the case that  $d_3 \in V(F, w_2)$ , and  $V(c_i) = d_i$ . As in every **Ct**-model,

$$V_{\mathcal{M}}(\exists x!F^\dagger(x), w, a) = T.$$

Now suppose for a reductio ad absurdum that the rule is sound. So,

$$V_{\mathcal{M}}(F^\dagger(\iota x F^\dagger(x)), w_1, a) = T.$$

By stipulation,  $V_{\mathcal{M}}(\neg F(c_3), w_1, a) = F$ . Consequently,

$$V_{\mathcal{M}}(\neg F(c_3) \wedge \Box \iota x F^\dagger(x) = c_2, w_1, a) = F.$$

Therefore,  $V_{\mathcal{M}}(\Box \iota x F^\dagger(x) = c_1, w_1, a) = T$ . This implies that

$$V_{\mathcal{M}}(\iota x F^\dagger(x) = c_1, w_2, a) = T.$$

If there is a unique  $F^\dagger$ , which is provably true, then the latter implies that  $V_{\mathcal{M}}(\exists x (F^\dagger(x) \wedge x = c_1), w_2, a) = T$ . Unpacking  $F^\dagger$  gives us that:

$$V_{\mathcal{M}}(\exists x (((F(c_3) \wedge x = c_1) \vee (\neg F(c_3) \wedge x = c_2)) \wedge x = c_1), w_2, a) = T.$$

The second disjunct cannot be true, because no intension can satisfy both  $x = c_1$  and  $x = c_2$ , since  $V(c_1) \neq V(c_2)$ . But the first disjunct is also false, because it is not the case that  $V(c_3) \in V(F, w_2)$ . Contradiction. This concludes the proof.  $\square$

Clearly, if a consequence, **W-UnrestrSelfPred**, of the description principle without the second restriction, is **C**-unsound, then the unrestricted description principle is **C**-invalid.

To sum up, **AtomSelfPred** and **NonModSelfPred** are **K+IC $\iota$** -valid, whereas **S-UnrestrSelfPred** and **W-UnrestrSelfPred** are respectively **Ct**-invalid and **Ct**-unsound. This teaches us that Carnap's restrictions were a good thing.

I have also claimed that self-predication principles can be useful in the discussion of the Føllesdal's reason for Carnap's first restriction: if one lifts the restriction, one can prove that modal distinctions collapse. In fact, Føllesdal has two collapse arguments [6, 69-72, 74]. Only the second one is directly relevant here, but to understand it, one is well-advised to first study his first collapse argument. That will be done in the next section.

## 5 Føllesdal's first collapse argument

In this section I will discuss Føllesdal's first collapse argument. In addition, I will also discuss a collapse argument by Quine that inspired Føllesdal's argument.

Føllesdal proved that, if one adds **NI** to classical Carnapian logic and makes one extra assumption, then one gets a collapse of necessity into truth. I will reformulate Føllesdal's theorem in  $\mathcal{L}^*$ .<sup>7</sup> Moreover, I will use **S5+ICt** as the background theory.

As was said earlier, Føllesdal's proof contains a derivation of an instance of a self-predication principle, namely **NonModSelfPred**. Let us make use of the general principle rather than deriving an instance of it. Here then is the modified theorem and proof:

**Theorem 5.1** (Føllesdal\*). *For all  $\phi \in \mathcal{L}^* \setminus \{\Box\}$ ,*

$$\forall x \forall y (x = y \rightarrow \Box x = y), \Box (z \neq c_0) \vdash_{\mathbf{S5+ICt}} \phi \rightarrow \Box \phi$$

*Proof.* Suppose that  $\phi \in \mathcal{L}^* \setminus \{\Box\}$  and that  $\Box z \neq c_0$ . By the law of self-identity and the introduction rule of conjunction one may infer that  $\phi \wedge z = z$ . By existential generalisation it follows that  $\exists x_{m+1} (\phi \wedge x_{m+1} = z)$ . The subscript of the variable has to do with the fact that there are at most  $m$  free variables in  $\phi$ , viz.  $x_1, \dots, x_m$ . By taking  $x_{m+1}$  one ensures that no free variable in  $\phi$  is bound by the existential quantifier. Also, make sure here not to replace any possible occurrences of  $z$  in  $\phi$ ! Transitivity of identity guarantees that  $\exists x_{m+1}! (\phi \wedge x_{m+1} = z)$  is then also inferable. Here **NonModSelfPred** can be used to get  $\iota x_{m+1} (\phi \wedge x_{m+1} = z) = z$ , because  $\phi$  was assumed to be a non-modal formula. Indeed,  $x_{m+1}$  is by stipulation not in  $\phi$  and  $x_{m+1} = z$  does not contain any modal operators. By **NI** and universal instantiation, one can infer that  $\Box \iota x_{m+1} (\phi \wedge x_{m+1} = z) = z$ . Let us abbreviate this as  $\Box \iota x_{m+1} \phi^* (x_{m+1}) = z$ . Next, it follows by **K $\iota$**  that

$$\Box (\exists x_{m+1} (!\phi^* (x_{m+1}) \wedge x_{m+1} = z) \vee (\neg \exists x_{m+1} !\phi^* (x_{m+1}) \wedge c_0 = z)).$$

It was assumed that  $\Box z \neq c_0$ . It follows in the modal system **K** that

$$\Box \exists x_{m+1} (!\phi^* (x_{m+1}) \wedge x_{m+1} = z)$$

Simplification leads to  $\Box \exists x_{m+1} \phi$ . It is a theorem of first-order logic that  $\exists x_{m+1} \phi \rightarrow \phi$  if  $x_{m+1}$  does not occur free in  $\phi$ , which is by the choice of  $x_{m+1}$  indeed the case. One can then use necessitation and the **K**-axiom to derive  $\Box \phi$  from  $\Box \exists x_{m+1} \phi$ .  $\square$

**Corollary 5.2.** *For all  $\phi \in \mathcal{L}^*$ ,*

$$\forall x \forall y (x = y \rightarrow \Box x = y), \Box (z \neq c_0) \vdash_{\mathbf{S5+ICt}} \phi \rightarrow \Box \phi$$

<sup>7</sup>Føllesdal used the contextual definition that corresponds to **C $\iota$** , whereas I will treat description terms as primitive and I will make use of **K $\iota$** . Moreover, Føllesdal made use of a proposition letter  $p$  in his proof, whereas there are no proposition letters in  $\mathcal{L}^*$ .

*Proof.* The proof is by induction on the complexity of  $\phi$ . One needs the **S5**-principle for the case in which  $\phi = \neg\psi$ .<sup>8</sup>  $\square$

**Corollary 5.3** (Føllesdal\*). *For all  $\phi \in \mathcal{L} \cup \{\iota\}$ ,*

$$\square\forall x\forall y(x = y \rightarrow \square x = y), z \neq c_0 \vdash_{\mathbf{S5+ICt}} \phi \rightarrow \square\phi$$

*Proof.* One needs the **B**-axiom to prove the necessity of difference from the necessity of **NI** ([6, p. 72], [7, p. 314]).  $\square$

It is easy to prove that **NI** is invalid in the aptly named *contingent identity* systems.<sup>9</sup>

Føllesdal's collapse argument was targeting Carnap's system of modal logic. It is related to a collapse argument given by Quine, who tried to argue for the claim that there are general difficulties with systems of modal logic. Quine's collapse argument goes as follows:

Schematically we can put the postulate as follows, using ' $Fx$ ' and ' $Gx$ ' (now) for arbitrary open sentences and using ' $Fx$  and  $x$  only' as short for ' $(w)$  ( $Fw$  if and only if  $w = x$ ):'

(4) If  $Fx$  and  $x$  only and  $Gx$  and  $x$  only then (necessarily  $(w)$  ( $Fw$  if and only if  $Gw$ ))

But this postulate annihilates modal distinctions; for we can deduce from it that 'Necessarily  $p$ ' holds no matter what true sentence we put for ' $p$ '. The argument is as follows. Let ' $p$ ' stand for any true sentence, let  $y$  be any object, and let  $x = y$ . Obviously then

(5)  $(p$  and  $x = y)$  and  $x$  only    and

(6)  $x = y$  and  $x$  only.

By (4), next, with its ' $Fx$ ' taken as  $(p$  and  $x = y)$  and its ' $Gx$ ' as ' $x = y$ ', we can conclude from (5) and (6) that

(7) Necessarily  $(w)((p$  and  $w = y)$  if and only if  $w = y)$ .

---

<sup>8</sup>Alternatively, one could have used a self-predication principle in which the descriptive condition is allowed to contain box operators, as long as the variable bound by the iota operator does not occur within the scope of any box operator.

<sup>9</sup>In [8, p. 249-250] it is proved that, if one enriches contingent identity systems with the Peano Arithmetic axioms, then **NI** is a theorem. Compare to [10, p. 249-250]. There are four main differences between the proof in [8] and the proof in [10]. First, in [10] the quantifiers are restricted to "Standard Numerals", viz.  $\forall x \in SN, \exists z \in SN$ , whereas in [8] they are not. Second, in [10]  $SN$  is defined as a second-order condition, viz.  $SN(x) =_{df} \forall X \forall z \{[(\square(z = 0) \rightarrow Xz) \wedge (\square Xz \rightarrow \square X(Sz))] \rightarrow Xx\}$ , whereas in [8] only first-order languages are considered. Third, in [10] the proof is carried out in second-order Peano Arithmetic and it involves class theoretical considerations (e.g., the minimal closure condition), whereas in [8] the proof is carried out in first-order Peano Arithmetic. Fourth, in [10] the proof of  $\forall x(\square x = \mathbf{0} \vee \exists y \square(x = \mathbf{s}(y)))$ , which is needed in the proof of the main result, the induction principle is used on top of the particular restriction of the quantifiers, whereas in [8] it is shown that the induction principle is sufficient. More about this result, induction principles and intensional mathematics can be found in [8].

But the quantification in (7) implies in particular ‘ $((p \text{ and } y = y) \text{ if and only if } y = y)$ ’, which in turn implies ‘ $p$ ’; so from (7) we conclude that necessarily  $p$ . [16, p. 197-198]

Perhaps the reader wonders where **NI** and the description principles come in here. The answer is that Quine’s line (4) is related to those principles. To make this more clear, I will make use of Martí’s partial reconstruction of Quine’s collapse argument, but I will also expand it.<sup>10</sup> Suppose that  $\phi$  is a true sentence and that  $c$  is a singular term denoting an object. (Quine used  $y$  but this is an insignificant difference.) Then  $\iota x (\phi \wedge x = c) = c$  is true. Now assume that the classical substitution principle is correct. Then one can get by universal instantiation the following:

$$\iota x (\phi \wedge x = c) = c \rightarrow (\Box c = c \rightarrow \Box \iota x (\phi \wedge x = c) = c)$$

Since  $\Box c = c$  is true, so must  $\Box \iota x (\phi \wedge x = c) = c$ . This mimics the crucial step in the proof of **NI** from the classical substitution principle. Martí says that  $\Box \iota x (\phi \wedge x = c) = c$  implies that  $\Box \phi$ . Although she does not explicitly say so, it is clear that Quine is relying here and in the first step on the necessity of the following description principle:

$$\mathbf{Q}_\iota \iota x (\phi \wedge x = c) = c \leftrightarrow \phi$$

To conclude, Quine used the principle of substitutivity of identity and a description principle to derive  $\phi \rightarrow \Box \phi$ . Quine’s (4) in the cited argument is the same as

$$\iota x (\phi \wedge x = c) = c \rightarrow \Box \iota x (\phi \wedge x = c) = c,$$

except for the fact that Quine uses the ‘expansions’ of the descriptions rather than the descriptions themselves (and for the insignificant difference that he uses the expansion of the description  $\iota x (x = c)$  rather than the term  $c$ ). As noted, Quine’s collapse argument was not intended to be tied to one particular system of modal logic. But one can of course investigate in which systems of modal logic it is a sound argument. In particular, one could look at the argument from a Carnapian perspective. Martí correctly noted that Quine’s collapse argument starts from an assumption, namely the classical substitution principle, that is alien to Carnap’s modal logic. Of course, one need not ascribe the belief that the classical substitution principle is sound under a Carnapian interpretation to Quine. The argument can be seen as detailing the consequences of adding that principle to Carnapian modal logic. But even then Quine’s argument is flawed.

$\mathbf{Q}_\iota$  is unsound under a Carnapian interpretation. For consider a world in which  $\phi$  is false. Let  $c$  refer to  $d^*$ . In that world  $\iota x (\phi \wedge x = c) = c$  is true but  $\phi$  is not. The invalidity of  $\mathbf{Q}_\iota$  was noted by [19] in connection with a so-called ‘slingshot argument’ given by Quine [17]. The argument goes as follows. Suppose that  $\Box \phi$  and  $\psi$  are true. By the **T**-axiom  $\phi$  is also true. Moreover, by Quine’s description principle  $\mathbf{Q}_\iota \iota x (\phi \wedge x = c) = c$  and  $\iota x (\psi \wedge x = c) = c$

---

<sup>10</sup>See [14, p. 280-281].

must be true. Hence,  $\iota x(\phi \wedge x = c) = \iota x(\psi \wedge x = c)$  is true by the transitivity of identity. Assuming that Quine's description principle is a necessary truth, it follows from  $\Box\phi$  that  $\Box\iota x(\phi \wedge x = c) = c$  is true. By the substitutivity of identicals,  $\Box\iota x(\psi \wedge x = c) = c$  is then also true. The latter implies  $\Box\psi$ . To conclude,  $(\Box\phi \wedge \psi) \rightarrow \Box\psi$ , or formulated in a slightly redundant way:  $(\phi \wedge \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$ .

There are three strong similarities between Quine's collapse argument (as reconstructed by Martí) and his slingshot argument. First, although the conclusions of Quine's collapse argument and his slingshot argument are different, this is only a superficial difference. The conclusion of Quine's collapse argument was that  $\phi \rightarrow \Box\phi$ . But if one substitutes the tautology  $\theta \vee \neg\theta$  for  $\phi$  in the conclusion of the slingshot argument, then the collapse statement follows. And the conclusion of the slingshot argument readily follows from the conclusion of the collapse argument. Second, in both arguments the classical substitution principle (or **NI**) is used. Third, in both arguments the description principle **Q $\iota$**  is used. Notwithstanding these obvious similarities Martí says the following:

[...] the specific collapse argument discussed here is not an instance of the slingshot, for it relies only on the assumption of the intersubstitution of codesignative terms,[footnote 14: This point is stressed in [Stephen Neale, 1995, On the philosophical significance of Gödel's slingshot, *Mind*]. See especially §14.] [...] [14, p. 282]

So she denies my third point and she resists my conclusion that these arguments are very similar.

As said above, Taylor noted that Quine's slingshot argument relied on a description elimination principle that is unsound under a Carnapian interpretation. He did not stop there. His solution was to add the assumption that  $c \neq \iota x(x \neq x)$ . Since everything is identical to itself,  $\iota x(x \neq x)$  refers to  $d^*$ . Now one can use a description elimination principle that is valid under a Carnapian interpretation:

$$(\iota x(\phi \wedge x = c) = c \wedge c \neq \iota x(x \neq x)) \rightarrow \phi \quad (1)$$

Except for the extra assumption and the different description elimination principle, Taylor's slingshot argument is exactly like Quine's. It is interesting to note that Taylor's slingshot argument is analogous to Føllesdal's collapse argument. Indeed, the latter worked with the assumption that  $c \neq c_0$  which is equivalent to the assumption that  $c \neq \iota x(x \neq x)$ . Moreover, Føllesdal derived in fact an instance of (1) (with  $c_0$  replacing  $\iota x(x \neq x)$ ). Taylor did not refer to Føllesdal, who constructed his collapse argument roughly twenty-four years earlier, so I assume that Taylor's work is independent.

In this section I have discussed Føllesdal theorem and its relation to collapse and slingshot arguments by Quine, Martí and Taylor. Føllesdal obtained a collapse result by assuming **NI**, but he knew only too well that the law of the substitutivity of identity did not hold in all systems of modal logic. In particular, it did not hold in the modal logic of Carnap. The latter replaced the classical law of substitutivity of identity by a principle stipulating substitutivity



of necessary identity ( $\equiv$  in Carnap’s symbolism). Føllesdal noted that, if one systematically replaced identity in his proof of by necessary identity, then one gets a collapse result even for Carnap’s system. So Føllesdal in fact obtained two collapse results for modal theories containing individual definite descriptions. Føllesdal’s second collapse argument will be discussed in the next section.

## 6 Føllesdal’s second collapse argument

Føllesdal’s second collapse argument relies essentially on an instantiation of **S-UnrestrSelfPred**.

**Theorem 6.1** (Føllesdal\*). *For all  $\phi \in \mathcal{L}^*$ ,*

$$\mathbf{S-UnrestrSelfPred}, \Box(z \neq c_0) \vdash_{\mathbf{T+CI}_t} \phi \rightarrow \Box\phi$$

*Proof.* Assume  $\phi$ . By the law of self-identity, necessitation and the introduction rule of conjunction, one can infer  $\phi \wedge \Box z = z$ . By existential generalisation, it follows that  $\exists x(\phi \wedge \Box x_{m+1} = z)$ . Transitivity of identity and the **T**-axiom warrant the derivability of  $\exists x!(\phi \wedge \Box x_{m+1} = z)$ . The application of **S-UnrestrSelfPred** results in  $\Box \iota x(\phi \wedge \Box x_{m+1} = z) = z$ . The remainder of the proof is as in the proof of theorem 5.1.  $\square$

Note that the above proof is given against the background of **T+CI<sub>t</sub>**, whereas the proof of Føllesdal’s first collapse argument was given against the background of the much less general theory **S5 + ICT**. One reason for this is that this time it was not necessary to use **GUI**.<sup>11</sup>

Although Føllesdal this time did not assume the invalid **NI**, he noted that Carnap restricted his description principle in such a way that the argument was still illegitimate:

Nevertheless the disastrous result in [theorem 5.1] is avoided in  $S_2$  when we use ‘ $\equiv$ ’ for the ‘ $=$ ’. For in the proof of [theorem 5.1], the description [ $\iota x(x \equiv z \wedge \phi)$ ] occurs in several of the lines. And since ‘ $x \equiv y$ ’ is short for ‘ $\Box(x = y)$ ’ (*Meaning and Necessity*, 39-6), descriptions of this type are not permitted by Carnap, who on p. 184 of *Meaning and Necessity* states that, “in order to avoid certain complications, which cannot be explained here, it seems advisable to admit in  $S_2$  only descriptions which do not contain ‘ $\Box$ ’.” Since in any proof of [theorem 5.1] we must be able to prove that [*phi*] is true if and only if the entity described in the description is necessarily distinct from [ $c_0$ ], it seems inevitable that at least one description in the proof contains the sign ‘ $\equiv$ ’, i.e., an  $\Box$ . [6, p.74]

Finally, one can understand why Føllesdal thought that Carnap imposed his first restriction (no box operators in the descriptive conditions). *Pace* Føllesdal,

<sup>11</sup>Another reason is that the self-predication principle is unrestricted and, hence, that there is no need to start from a non-modal truth, so that there is no need to invoke the **S4**-axiom to generalize the result.

Martí claims that Føllesdal's second collapse argument fails, not because it violates Carnap's first restriction, but because it violates his second restriction [13, p. 580-583]. The step from  $\exists x!(\phi \wedge \Box x_{m+1} = z)$  to  $\Box \iota x(\phi \wedge \Box x_{m+1} = z) = z$  is where the argument breaks down. This step is based on Carnap's / Kremer's description principle as follows: if the first has been derived, then one can also derive  $\exists x!(\phi \wedge \Box x_{m+1} = z) \wedge \Box x_{m+1} = z$ . From there one can derive the Carnap/Kremer disjunctive expansion, and then one can use the description principle to get  $\Box \iota x(\phi \wedge \Box x_{m+1} = z) = z$ . But the description principle was restricted so as to allow only primitive predicates or the identity predicate to be used to predicate something of the *descriptum*, and a necessary identity is a complex predicate. So Carnap's second restriction is violated. The effect of this violation is that one can infer  $\Box \exists x!(\phi \wedge x_{m+1} = z)$  from  $\exists x!(\phi \wedge \Box x_{m+1} = z)$ . And that is surely an invalid inference.

Sofar I have only argued that Føllesdal's second collapse argument violates Carnap's second restriction and that the violation yields an invalid inference. But the argument also violates Carnap's first restriction, since the descriptive condition contains a box operator. Martí shows that the latter is really inessential. She does so by showing that one can construct a variation on Føllesdal's collapse argument that respects Carnap's first restriction but still violates his second restriction [13, app. II]. The idea is to prove  $\Box \iota x_{m+1}(x_{m+1} = z \wedge \phi) = z$  from

$$\exists x_{m+1}(! (x_{m+1} = z \wedge \phi) \wedge \Box x_{m+1} = z),$$

with  $\phi$  being non-modal. I will not provide the details, because it is easy enough to reconstruct her argument on the basis of my presentation of Føllesdal's collapse argument. The only difference consists in the fact that one will need not the unrestricted self-predication principle as such, but a principle that is very close to one (and also follows from the unrestricted  $\mathbf{K}\iota$  principle).

In my presentation of Føllesdal's second collapse argument I used an application of **S-UnrestrSelfPred**. It was already shown that this principle is invalid. **NonModSelfPred** is valid however. According to this presentation, what is going wrong in the collapse argument is not that a complex predicate is being used, but that the complex predicate contains a box operator. This point can be strengthened by noting that, if  $\psi$  is non-modal, then one can validly infer  $\psi(\iota x\phi(x))$  from  $\exists x(!\phi(x) \wedge \psi(x))$ . (This is so for basically the same reason that the self-predication principle restricted to non-modal formulas is valid.)

One might wonder why Carnap imposed his second restriction. One could ask why he did not use lambda terms  $\lambda$  and then lay down the following rule:

$$(\lambda y\psi(y))(\iota x\phi(x)) \leftrightarrow \exists x(!\phi(x) \wedge \psi(x)) \vee (\neg \exists x!\phi(x) \wedge \psi(c_0))$$

However, Martí notes that:

There is nothing wrong with the (apparently *de re*) sentence  $[\lambda z\Box Qz]\iota xPx$ ; it so happens though that in S2 that sentence is equivalent to the obviously *de dicto*  $\Box[\lambda zQz]\iota xPx$ : "For lambda-expressions we do not impose the restriction stated for descriptions; they may also contain [' $\Box$ ']. Any lambda operator can be eliminated in S2 by conversion ..." and Carnap

proceeds then to mention as the rule of lambda conversion for S2 the same unrestricted rule that he had introduced for S1, the extensional predecessor of S2. So, the apparently *de re*  $[\lambda z \Box Qz] \iota x P x$  is not ruled out of S2 by some unjustified restriction. It is a perfectly well formed sentence, but it is not significantly different from its *de dicto* counterpart.

[13, p. 582]

The claim made in the first part of Martí's quote does not seem to be warranted, because the lambda conversion rule of S1 (Carnap's non-modal logic with iota and lambda terms) is the following:

If a sentence consists of an abstract expression followed by an individual constant, it says that the individual has the property in question. Therefore,  $[(\lambda x)(\dots x \dots) c]$  means the same as  $[\dots c \dots]$ , that is, the sentence formed from  $[\dots x \dots]$  by substituting  $[c]$  for  $[x]$ . The rules of our system will permit the transformation of  $[(\lambda x)(\dots x \dots) c]$  into  $[\dots c \dots]$  and vice versa; these transformations are called conversions. [3, p. 3]

It is clear that, on the basis of the  $\lambda$ -conversion rule given by Carnap,

$$(\lambda y \Box \psi(y)) c$$

and

$$\Box (\lambda y \psi(y)) c$$

are equivalent. But Carnap's conversion rule is for combinations of lambda terms with individual constants (and probably also for individual variables), but the rule does not say anything about combinations of lambda terms with iota terms. If the conversion rule for  $(\lambda y \psi(y)) t$  were analogous to the conversion rule for  $(\lambda y \psi(y)) c$ , then Martí would be right. Perhaps she still is right, but not for the reason cited. Moreover, if one had the conversion rule for combinations of lambda and iota terms, then one can understand the rationale behind Carnap's second restriction. Martí gives an interesting reason for the claim that Carnap *ought* to have had the expanded conversion rule, namely his anti-essentialism should prohibit him attributing *de re* properties to individuals. This is an interesting point that is worth exploring further, but that will not be done in this paper.

To conclude this section, I present a collapse argument that also involves the unrestricted self-predication principle. First I need to prove a lemma. Then I will prove the main theorem.

**Lemma 6.2.** *For all  $\phi \in \mathcal{L}^*$ , if*

$$\phi^* =_{df} ((\phi \leftrightarrow \Box x = z_1) \wedge (\neg \phi \leftrightarrow \Box x = z_2)) \wedge (\Box \Box x = z_1 \vee \Box \Box x = z_2),$$

*then  $\Diamond z_1 \neq z_2 \vdash_{\mathbf{T+CI}} \exists x! \phi^{**}(x)$*

*Proof.* I will first prove the existence claim and then the uniqueness claim.

*Existence.* Assume that  $\diamond z_1 \neq z_2$ . The following is a tautological consequence of the assumption, the law of self-identity and the rule of necessitation:

$$\begin{aligned} & ((\phi \leftrightarrow \Box z_1 = z_1) \wedge (\neg\phi \leftrightarrow \Box z_1 = z_2)) \wedge (\Box\Box z_1 = z_1 \vee \Box\Box z_1 = z_2) \quad (2) \\ & \vee ((\phi \leftrightarrow \Box z_2 = z_1) \wedge (\neg\phi \leftrightarrow \Box z_2 = z_2)) \wedge (\Box\Box z_2 = z_1 \vee \Box\Box z_2 = z_2) \end{aligned}$$

Existential generalisation on the first disjunct with  $x$  systematically replacing  $z_1$ , existential generalisation on the second disjunct with  $x$  systematically replacing  $z_2$ , and distributing the existential quantifiers over the whole disjunction results in  $\exists x\phi^*(x)$ .

*Uniqueness.* Suppose that  $\forall x(\neg\phi^*(x) \vee \exists y(\phi^*(y) \wedge y \neq x))$ . We had already established that  $\exists x\phi^*(x)$ . Suppose that  $\phi^*(t)$  for some term  $t$ . By universal instantiation, one also has  $\neg\phi^*(t) \vee \exists y(\phi^*(y) \wedge y \neq t)$ . The first disjunct can be eliminated, since it conflicts with the assumption. Suppose then that  $\phi^*(t') \wedge t \neq t'$  for some term  $t'$ . Assume that  $\phi$ . It is a tautological consequence of the previous and the **T**-axiom that  $t = z_1 \wedge t' = z_1 \wedge t \neq t'$ . The latter contradicts the principle of the transitivity of identity. One can run an analogous argument starting from the assumption that  $\neg\phi$ . Ergo, by constructive dilemma and existential instantiation, contradiction.  $\square$

**Theorem 6.3.** *For all  $\phi \in \mathcal{L}^*$ ,*

$$\Box\mathbf{S}\text{-UnrestrSelfPred}, \Box\diamond z_1 \neq z_2 \vdash_{\mathbf{T}+\mathbf{CI}} \phi \rightarrow \Box\phi$$

*Proof.* By lemma 6.2, necessitation, the **K**-axiom and the assumption that  $\Box\diamond z_1 \neq z_2$ , it follows that  $\Box\exists x!\phi^*(x)$ , with  $\phi^*(x)$  being defined as in the proof of lemma 6.2. By the necessitation of **S-UnrestrSelfPred**, the **K**-axiom and the **T**-axiom, it follows that  $(\Box\phi \leftrightarrow \Box\Box\iota x\phi^*(x) = z_1)$  is true, that  $(\Box\neg\phi \leftrightarrow \Box\Box\iota x\phi^*(x) = z_2)$  is true, and that  $\Box\Box\iota x\phi^*(x) = z_1 \vee \Box\Box\iota x\phi^*(x) = z_2$  is true. It is a tautological consequence that  $\Box\phi \vee \Box\neg\phi$ . By the **T**-axiom, it follows that  $\phi \rightarrow \Box\phi$ .  $\square$

The above collapse result is interesting for two reasons.<sup>12</sup> First, unlike the other collapse and slingshot arguments discussed in this paper, it does not make use of a description elimination principle, e.g. the left-to-right direction of **C $\iota$** , **K $\iota$** , or **Q $\iota$** . This is another reason for focussing on self-predication principles: they are a special kind of description *introduction* principles and, in order to prove a collapse result, one does not need any description *elimination* principle. Second, although the above collapse argument uses one stronger assumptions ( $\Box\mathbf{S}\text{-UnrestrSelfPred}$ ) than Føllesdal's second collapse argument (**S-UnrestrSelfPred**), it also uses a weaker assumption ( $\Box\diamond z_1 \neq z_2$ ) than

<sup>12</sup>A variation on this argument has also a philosophical application: see [9]. There the background theory was some form of intensional mathematics, in which  $\Box\mathbf{s}(\mathbf{0}) \neq \mathbf{0}$  is a theorem. Also, another self-predication principle has been used. Instead of allowing unrestricted self-predication on the condition of the satisfaction of the existence and uniqueness clause, the other principle allows unrestricted self-predication on the condition of the *necessary* satisfaction of the existence and uniqueness condition.

Føllesdal's collapse argument ( $\Box z \neq c_0$ ), and it uses no other assumptions than Føllesdal's.

Føllesdal's second collapse argument was a variation on his first collapse argument, which used **NI**. Clearly, there is a variation on the above collapse argument, which uses **S-UnrestrSelfPred**, that starts from the assumption of **NI** and the necessity of **NonModSelfPred**. Of course, in that case one would also need **GUI**, which one does not need in the above collapse argument. In fact, one does not need any kind of universal instantiation principle for description terms.

To conclude, Føllesdal has a second collapse argument that involves an instantiation of an unrestricted self-predication principle, that is invalid. Martí claims rightly that his argument violates Carnap's second restriction, but it is still not entirely clear why Carnap did impose that restriction in the first place. But is clear that, even when one drops Carnap's restriction and adopts a more relaxed restriction, that Føllesdal's second collapse argument goes wrong. Finally, I have shown there is a collapse argument that also involves an instantiation of an unrestricted self-predication principle and that is weaker in some respects than Føllesdal's.

## 7 Conclusion

In this paper I have discussed Carnapian modal logic with descriptions. Carnap imposed two restrictions on his description principle: first, the descriptive conditions of the description terms should be non-modal; second, only primitive predicates or the identity predicate can be used to predicate something of the *descriptum*. Martí argued that, if one formulates the existence and uniqueness conditions as did Carnap, then for a broad class of modal descriptions the existence and uniqueness conditions are never satisfied. I argued that this problem is limited to **C** and that it only partially justifies Carnap's first restriction. She then proposed a solution, which was shown to be inadequate. Kremer proposed another solution, which does the job it was designed for. Carnap's first restriction becomes obsolete and the problem disappears.

Føllesdal had a different theory about the reason for Carnap's first restriction. He thought that, if one lifts the restriction, then one gets a collapse of necessary truth into plain truth. He gave two collapse arguments. The first collapse argument starts from the principle of the necessity of identity, which is invalid in a Carnapian interpretation. It was argued that there is an argument to be found in the literature on slingshot arguments that is very similar to Føllesdal's first collapse argument. To remedy his first collapse argument, Føllesdal gave a second collapse argument. Martí claimed, rightly, that Føllesdal's second collapse argument violated Carnap's second restriction, and essentially only that restriction. She also gave a reason Carnap might have had for imposing this restriction, namely his lambda-conversion rules, but it was pointed out that the text does not support that claim. She also presented a second rationale, but that one has not been critically discussed.

Throughout the paper I have made use of self-predication principles. They were used to argue that Carnapian theories of description deal inadequately with improper descriptions. They were also used to make the structure of Føllesdal's arguments more transparent and to pinpoint where exactly his second collapse argument goes wrong. Finally, I have given a collapse argument that starts from assumptions that are in some respects weaker than Føllesdal's.

**Acknowledgements.** Research for this paper was partially supported by a grant of the Fund for Scientific research–Flanders, which is gratefully acknowledged. I would like to thank the editor and two anonymous referees for their many good suggestions and comments. Thanks especially to Leon Horsten for his many comments on the series of papers leading up to this paper.

**Author** Jan Heylen  
Institute of Philosophy, K.U.Leuven  
Dekenstraat 2, 3000 Leuven, Belgium  
jan.heylen@hiw.kuleuven.be

## References

- [1] George S. Boolos, John P. Burgess, and Richard C. Jeffrey. *Computability and Logic*. Cambridge University Press, New York, fourth edition, 2003.
- [2] Rudolf Carnap. Modalities and quantification. *Journal of Symbolic Logic*, 11(2):33–64, 1496.
- [3] Rudolf Carnap. *Meaning and Necessity*. The University of Chicago Press, Chicago, second edition, 1956.
- [4] Rudolf Carnap. *The Philosophy of Rudolf Carnap*, volume XI of *The Library of Living Philosophers*, chapter My Conceptions of the Logic of Modalities, pages 889–900. Open Court, La Salle, Illinois, 1963.
- [5] Dagfinn Føllesdal. Quine on modality. *Synthese*, 19(1-2):147–157, 1968.
- [6] Dagfinn Føllesdal. *Referential Opacity and Modal Logic*. Routledge, 2004. Originally published as: *Referential Opacity and Modal logic*. Oslo: Universitetsforlaget, 1966.
- [7] M. Cresswell G. Hughes. *A New Introduction to Modal Logic*. Routledge, London, 1996.
- [8] Jan Heylen. Carnapian modal and epistemic arithmetic. In Massimiliano Carrara and Vittorio Morato, editors, *Language, Knowledge, and Metaphysics. Proceedings of the First SIFA Graduate Conference*, pages 97–121. College Publications, London, 2008.
- [9] Jan Heylen. Descriptions and unknowability. *Analysis*, 70(10):50–52, 2010.

- [10] Leon Horsten. Canonical naming systems. *Minds and Machines*, 15(2):229–257, 2005.
- [11] Michael Kremer. Marti on descriptions in Carnap’s s2. *Journal of Philosophical Logic*, 26(6):629–634, 1997.
- [12] Saul A. Kripke. Individual concepts: Their logic, philosophy and some of their uses. In *Proceedings and Addresses of the American Philosophy Association*, volume 66, pages 70–73, 1992.
- [13] Genoveva Martí. Do modal distinctions collapse in Carnap’s system? *Journal of Philosophical Logic*, 23(6):575–593, 1994.
- [14] Genoveva Martí. Rethinking Quine’s argument on the collapse of modal distinctions. *Notre Dame Journal of Formal Logic*, 38(2):276–294, 1997.
- [15] Terence Parsons. Essentialism and quantified modal logic. *The Philosophical Review*, 78(1):35–52, 1969.
- [16] Willard Van Orman Quine. *Word and Object*. M.I.T. Press, 1960.
- [17] Willard Van Orman Quine. *The Ways of Paradox and Other Essays*, chapter Three Grades of Modal Involvement, pages 156–174. Random House, New York, 1966.
- [18] Gerhard Schurz. Carnap’s modal logic. In W. Stelzner and M. Stöckler, editors, *Zwischen traditioneller und moderner Logik. Nichtklassische Ansätze*, pages 365–380. Mentis Verlag, 2001.
- [19] B. Taylor. *Modes of Occurrence. Verbs, Adverbs, and Events*. Blackwell, Oxford, 1985.