Carnapian Modal and Epistemic Arithmetic*

Jan Heylen

The subject of the first section is Carnapian modal logic. One of the things I will do there is to prove that certain description principles, viz. the "self-predication principles", i.e. the principles according to which a descriptive term satisfies its own descriptive condition, are theorems and that others are not. The second section will be devoted to Carnapian modal arithmetic. I will prove that, if the arithmetical theory contains the standard weak principle of induction, modal truth collapses to truth. Then I will propose a different formulation of Carnapian modal arithmetic and establish that it is free of collapse. Noteworthy is that one can retain the standard strong principle of induction. I will occupy myself in the third section with Carnapian epistemic logic and arithmetic. Here too it is claimed that the standard weak principle of induction is invalid and that the alternative principle *is* valid. In the fourth and last section I will get back to the self-predication principles and I will point to some of the consequences if one adds them to Carnapian Epistemic arithmetic. The interaction of self-predication principles and the strong principle of induction results in a collapse of *de re* knowability.

1 Carnapian logic

First, I will say a few words about the language and quite a few more words on the formal and informal interpretation of the language. Next, I will then briefly sketch the theory and discuss the formal and informal soundness of the principles. Finally, I will prove and disprove a few description principles.

1.1 The language and its interpretation

The non-logical terminology of $\mathcal{L}$ consists of a set of individual constants, $\mathcal{C}_{\mathcal{L}}$, a distinguished constant $c^* \in \mathcal{C}_{\mathcal{L}}$, a set of functions, $\mathcal{F}_{\mathcal{L}}$, and a set of predicates, $\mathcal{P}_{\mathcal{L}}$. Sofar the only unusual feature is c^*. Its role and importance will become clearer later.

The set of well-formed terms and formulas can be defined recursively as follows:

Definition 1 For all $c_i \in \mathcal{C}_{\mathcal{L}}$, c_i is a well-formed term; if $t_1, \ldots, t_n$ are well-formed terms and if $f_i^n \in \mathcal{F}_{\mathcal{L}}$, then $f_i^n(t_1, \ldots, t_n)$ is a well-formed term; if ϕ is a well-formed formula, then $\iota x\phi$ is a well-formed term; if $t_1, \ldots, t_n$ and if $P^n \in \mathcal{R}_{\mathcal{L}}$, then $t_i = t_j$ (with $1 \leq i \leq j \leq n$) and $P^n(t_1, \ldots, t_n)$ are well-formed formulas; if ϕ, ψ are well-formed formulas, then $\neg\phi, \phi \wedge \psi, \phi \vee \psi, \phi \rightarrow \psi, \phi \leftrightarrow \psi, \Box\phi$ are well-formed formulas; if ϕ is a formula with, then $\forall x\phi$ and $\exists x\phi$ are well-formed formulas; nothing else is a well-formed term or formula. □

*An earlier version of this paper was presented at the First SIFA Graduate Conference (Padua, 11-12 September 2007). A later version was presented at a seminar organised by the CLAP (University of Leuven). The final version was presented at the Sixth ESAP Conference (Krakau, 21-26 August 2008). Research for this paper was supported by a grant of the Fund for Scientific research –Flanders, which is gratefully acknowledged. I would like to thank the audiences at those events for their comments. Thanks especially to Leon Horsten for his many comments on the series of papers leading up to and including this paper.

The set of closed terms, $\tau_{\mathcal{L}}$, contains all well-formed terms in which no variable is free. The set of closed formulas, $\mathcal{S}_{\mathcal{L}}$, contains all well-formed formulas in which no variable is free.

An unusual feature of the above definition is that the iota-operator is conceived as a term-forming operator and not as a formula-forming operator. In this respect the ι-operator differs from the quantifier $\forall$. Both operate on a formula ϕ, but the first transforms ϕ in an term, whereas the latter transforms ϕ in a new formula. On an alternative conception the iota-operator should also be a formula-forming operator. On that line of approach descriptions are like quantifiers. $(\iota x \phi(x))(\psi(x))$ should be read as: "concerning the ϕ, ψ". Russell's theory of descriptions was in a way a mixture of the term approach on the one hand and the quantifier approach on the other hand, since he allowed descriptions as quasi-terms but defined away formulas with descriptions occurring in them as complex quantified formulas.

Usually one gives separate definitions of the set of well-formed terms on the one hand and the set of well-formed formulas. The choice for descriptions as terms blocks this approach however, for the separate definitions would then be circular, since one step in the definition of the well-formed terms requires a well-defined set of well-formed formulas (viz. the recursive step in which ι-terms are introduced), while one step in the definition of well-formed formulas requires that one has a well-defined set of well-formed terms (viz. the basic step in which the atomic formulas are listed as well-formed formulas). By giving one recursive definition for both the well-formed terms as the well-formed formulas one avoids circularity. First, one builds all description-free well-formed terms. Next, one uses these building blocks to build all description-free well-formed terms. Then one can build the description terms the descriptive condition of which does not contain other description terms. Subsequently one can use those description terms to construct the well-formed formulas which contain at most the description terms which are already defined, and so on.

It was already pointed out that the set of well-formed terms and the set of well-formed formulas have to be defined simultaneously. As a consequence, the interpretation of the terms and formulas also need to be given in one definition:

Definition 2 A model is a quintuple $\langle W, @, D, d^*, I\rangle$ where W and D are non-empty sets, $@ \in W$, $d^* \in D$ and I is a function defined as follows: if $c \in \mathcal{C}_{\mathcal{L}}$, then $I(c) \in D$; $I(c^*) = d^*$; if $P \in \mathcal{P}_{\mathcal{L}}$, then for all $w \in W$, $I(P, w) \subset D^n$. Let i be a function from W to D and let a be a function such that, if x is a variable, $a(x) = i$ for some i. Let $a[i/x]$ be a function differing from a at most in mapping i to x. Then let $V_{\mathcal{M}}$ be a function defined as follows: if x is a variabele, $V_{\mathcal{M}}(x, w, a) = (a(x))(w)$; if $c \in \mathcal{C}_{\mathcal{L}}$, $V_{\mathcal{M}}(c, w, a) = I(c)$; if $P^n \in \mathcal{P}_{\mathcal{L}}$, $V_{\mathcal{M}}(P(t_1, \ldots, t_n), w, a) = T$ iff $\langle V_{\mathcal{M}}(t_1, w, a), \ldots, V_{\mathcal{M}}(t_n, w, a)\rangle \in I(P, w)$; $V_{\mathcal{M}}(t_1 = t_n, w, a) = T$ iff $V_{\mathcal{M}}(t_1, w, a) = V_{\mathcal{M}}(t_2, w, a)$; the clauses for $\neg$ and $\rightarrow$ 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; $V_{\mathcal{M}}(\Box \phi, w, a) = T$ iff $V_{\mathcal{M}}(\phi, w', a) = T$ for every $w' \in W$; $V_{\mathcal{M}}(\iota x \phi, w, a) = d$ if

1. for some i, $V_{\mathcal{M}}(\phi, w, a[i/x]) = T$ and
2. 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. $\square$

If ϕ is closed, then $V_{\mathcal{M}}(\phi, w) = V_M(\phi, w, a)$. Finally, let

$$V_{\mathcal{M}}(\phi) = V_M(\phi, @).$$

The above formal interpretation of $\mathcal{L}$ was given by Kremer (1997, p. 632-64).

1.2 The theory and its soundness

Carnapian modal logic, **C**, is the combination of first-order logic, the modal system **S5** and some characteristic principles regarding identity and descriptions. In the paper I will make use of a weaker theory, $\mathbf{C_T}$, which is exactly like **C**, except for the fact that it comprimes the modal system **T** instead of **S5**. The sole motivation for doing so is, of course, that the weaker the theory, the stronger the proof of the theoremhood of a certain formula.

First-order logic is well-known, and need not be described in full detail here. Suffice it to say that the quantifier principles are the following:[1]

∀E If $\Gamma \vdash \forall x\phi$, $\Gamma \vdash \phi[x|t]$, provided that t is free for x in ϕ.

∀I If $\Gamma \vdash \phi$ and x does not occur free in any member of Γ, then $\Gamma \vdash \forall x\phi$.

Modal system **T** consists of two rules:[2]

□E If $\Gamma \vdash \Box\phi$, then $\Gamma \vdash \phi$

□I-T If $\Gamma \vdash \phi$, then $(\Box)\,\Gamma \vdash \Box\phi$, with $(\Box)\,\Gamma = \{\Box\psi|\psi \in \Gamma\}$

The identity principles are somewhat deviant. There are two elimination rules for identity rather than one.[3]

=E If $\Gamma_1 \vdash t_1 = t_2$ and if $\Gamma_2 \vdash \phi$ with no occurrences of $\Box$ in ϕ, then $\Gamma_1, \Gamma_2 \vdash \phi'$ where ϕ' is obtained from ϕ by replacing zero or more occurrences of t_1 with t_2, provided that no bound variables are replaced, and if t_2 is a variable, then all of its substituted occurrences are free.

□ =E If $\Gamma_1 \vdash \Box t_1 = t_2$ and if $\Gamma_2 \vdash \phi$, then $\Gamma_1, \Gamma_2 \vdash \phi'$ where ϕ' is obtained from ϕ by replacing zero or more occurrences of t_1 with t_2, provided that no bound variables are replaced, and if t_2 is a variable, then all of its substituted occurrences are free.

The deviancy lies in the fact that =**E** is usually taken be holding of *all* well-formed formulas, not merely of all the box-free well-formed formulas. For the latter □ =**E** is needed. The only I would like to note here is that it is a good thing that =**E** is restricted to non-modal formulas. It is well-known that, if $\forall x \Box x = x$ is a theorem, which it is in $\mathbf{C_{S5}}$, and if one has the unrestricted =**E**-rule, then one can prove that $\forall x\,(x = y \rightarrow \Box x = y)$. This formula will play a role later on. So the elimination rule for identity is deviant, but the introduction rule for identity on the other hand is non-deviant.

=I $\Gamma \vdash t = t$, where t is any term.

When formulating the rules for descriptions, it will be convenient to use the following abbreviatory definition:

Definition 3 $!\phi\,(t) =_{df} \phi\,(t) \wedge \forall y\,(\phi\,(y) \rightarrow y = t)$ □

The introduction and elimination rules for descriptions can then be formulated as follows:

ιI If $\Gamma \vdash \psi\,(\iota x\phi\,(x))$, then $\Gamma \vdash \exists x\,(!\phi\,(x) \wedge \psi\,(x)) \vee (\neg\exists x!\phi\,(x) \wedge \psi\,(c^*))$.

[1]See Shapiro (2007, section 3).
[2]See G. Hughes (1996, p. 214, 216)
[3]For =**E** see Shapiro (2007, section 3)

ιE If $\Gamma \vdash \exists x\left(!\phi\left(x\right) \wedge \psi\left(x\right)\right) \vee \left(\neg\exists x!\phi\left(x\right) \wedge \psi\left(c^{*}\right)\right)$, then $\Gamma \vdash \psi\left(\iota x\phi\left(x\right)\right)$.

These rules are the natural translation of a contextual definition given by Kremer (1997, p. 631) as an improvement on the definition given by Carnap himself and an earlier improvement given by Marti (1994). I will not go into the whys. There is a very important restriction on the last contextual definition: descriptions should always be assigned the smallest possible scope. E.g., if ψ is of the form $\Box\varphi$ and if φ is a atomic well-formed formula, then the definiens of $\Box\varphi\left(\iota x\phi\left(x\right)\right)$ is the following:

$$\Box\left(\exists x\left(!\phi\left(x\right) \wedge \varphi\left(x\right)\right) \vee \left(\neg\exists x!\phi\left(x\right) \wedge \varphi\left(c^{*}\right)\right)\right)$$

This concludes the description of the theory.

Both **C** and $\mathbf{C}_T$ are sound under the Kremer interpretation, as the reader can verify for him or herself. If I had added a reflexive accessibility relation to the models, one would get a class of models which validate $\mathbf{C}_T$ but not **C**. I have not done so, however, since the more encompassing models are, the stronger the proof of the non-theoremhood of a certain formula.

1.3 Self-predication principles

It is intuitively true that, if there is one president of France, then the president of France is a president (of France). Let us call this description principle the self-predication principle. So it is a good thing that the following is a theorem of $\mathbf{C_T}$:

Theorem 1 $\vdash_{\mathbf{C_T}} \exists x!\phi\left(x\right) \rightarrow \phi\left(\iota x\phi\left(x\right)\right)$ *for all box-free* ϕ □

PROOF Assume that $\exists x!\phi\left(x\right)$. Suppose that $!\phi\left(t\right)$ for some term t, i.e. suppose that $\phi\left(t\right) \wedge \forall y\left(\phi\left(y\right) \rightarrow y = t\right)$. By =**I** and $\wedge$**I** one can infer from the previous that $\phi\left(t\right) \wedge \forall y\left(\phi\left(y\right) \rightarrow y = t\right) \wedge t = t$. Next, by $\exists$**I** one may infer that $\exists x\left(\phi\left(x\right) \wedge \forall y\left(\phi\left(y\right) \rightarrow y = x\right) \wedge x = t\right)$. Application of $\vee$**I** delivers one then the definiens of $t = \iota x\phi\left(x\right)$ and, therefore, also the definiendum. If ϕ is box-free, one may use =**E** to infer from $t = \iota x\phi\left(x\right)$ and $\phi\left(t\right)$ that $\phi\left(\iota x\phi\left(x\right)\right)$. Finally, it follows by $\exists$**E** from $\exists x!\phi\left(x\right)$ that $\phi\left(\iota x\phi\left(x\right)\right)$. ■

The above self-predication principle is restricted to non-modal well-formed formulas. One might wonder if it is possible to generalize the self-predication principle to all well-formed formulas, be they non-modal or modal. The answer is that it is not possible.

Theorem 2 *There is a modal well-formed formula* ϕ *and a Carnapian model of* $\mathbf{C_T}$ *in which* $\exists x!\phi\left(x\right)$ *is true but* $\phi\left(\iota x\phi\left(x\right)\right)$ *is not true.* □

PROOF Let us take $\Box F\left(x\right)$ as an example of a modal well-formed formula and consider the following model: $D = \{d_1, d_2, d^{*}\}$; $W = \{@, w_1\}$; R is a total two-place relation on W; $I\left(F, @\right) = \{d_1\}$; $I\left(F, w_1\right) = \{d_1, d_2\}$. Next, consider the following individual concept: $i_1 \mapsto d_1$.

Clearly, $i_1\left(w'\right) \in I\left(F, w'\right)$ for all $w' \in W$. It follows that

$$V_{\mathcal{M}}\left(\Box F\left(x\right), @, a[i_1/x]\right) = T.$$

Next thing to verify is that for all i', if $V_{\mathcal{M}}\left(\Box F\left(y\right), @, a[i'/y]\right) = T$, then $i'\left(@\right) = i_1\left(@\right) = d_1$. There are three cases to be considered. First, if $i'\left(@\right) = d_1$, then the consequent is true and, therefore, the conditional is true. Second, if $i'\left(@\right) = d_2$, then $V_{\mathcal{M}}\left(\Box F\left(y\right), @, a[i'/y]\right) = F$, since it is not the case that $d_2 \in I\left(F, @\right)$ and since R is a total relation on W. Third, if $i'\left(@\right) = d^{*}$, then $V_{\mathcal{M}}\left(\Box F\left(y\right), @, a[i'/y]\right) = F$, for the same reason as in the previous case. In conclusion, $V_{\mathcal{M}}\left(\exists x!\Box F\left(x\right), @, a\right) = T$.

This leaves us with checking that $V_{\mathcal{M}}\left(\Box F\left(\iota x\Box F\left(x\right)\right),@,a\right)=F$. The latter is false if and only if $V_{\mathcal{M}}\left(\iota x\Box F\left(x\right),@,a\right)\in I\left(F,@\right)$, and

$$V_{\mathcal{M}}\left(\iota x\Box F\left(x\right),w_{1},a\right)\in I\left(F,w_{1}\right).$$

The first condition is fulfilled, since $d_1\in\{d_1\}$. The latter condition is satisfied if and only if either for some i,

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

and for every i, if $V_{\mathcal{M}}\left(\Box F\left(x\right),w,a[i/x]\right)=T$, then $i\left(w\right)=d_k$ with $k\in\{1,2\}$, and $d_k\in I\left(F,w_1\right)$, or $d^*\in I\left(F,w_1\right)$. Let us consider the disjuncts in order of presentation. Next to the earlier introduced i_1 there are only three other individual concepts: $i_2=_{df}\ i_2\left(@\right)=d_1;i_2\left(w_1\right)=d_2$; $i_3=_{df}\ i_2\left(@\right)=d_2;i_2\left(w_1\right)=d_1$; $i_4=_{df}\ w\mapsto d_2$. The last two individual concepts will not do, since $V_{\mathcal{M}}\left(\Box F\left(x\right),w_1,a[i_l/x]\right)=F$ with $l\in\{3,4\}$, because it is not the case that $d_2\in I\left(F,@\right)$. The first two individual concepts can pass this barrier however. But the fact that they denote two different individuals in w_1 barrs them from taking the last hurdle. So the first disjunct can be dismissed. This leaves us with the second disjunct. Since it is not the case that $d^*\in I\left(F,w_1\right)$, one may dismiss the second disjunct too. Therefore the second condition is not satisfied and $V_{\mathcal{M}}\left(\Box F\left(\iota x\Box F\left(x\right)\right),@,a\right)=F$ is consequently true. ■

So the self-predication principle cannot be generalized to all well-formed formulas while at the same time retaining soundness in $\mathbf{C_T}$-models. But one might ask, what if considers weaker self-predication principles which hold for all well-formed formulas nonetheless? One possibility is that the antecedent of the self-predication principle ought not to be merely true but also necessarily true. Unfortunately, this is no good either.

Theorem 3 *There is a modal well-formed formula ϕ and there is a $\mathbf{C_T}$-model in which $\Box\exists x!\phi\left(x\right)$ is true but $\phi\left(\iota x\phi\left(x\right)\right)$ is not true.* □

Proof Take $F\left(x\right)\wedge\neg\Box F\left(x\right)$ as an example of a modal well-formed formula and consider the following model: $D=\{d_1,d_2,d^*\}$; R is a total reflexive two-place relation on D; $W=\{@,w_1\}$; $I\left(F,@\right)=\{d_1\}$; $I\left(F,w_2\right)=\{d_2\}$. The model makes it true that

$$V_{\mathcal{M}}\left(\Box\exists x!\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right),@,a\right).$$

Indeed, one has both

$$V_{\mathcal{M}}\left(\exists x!\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right),@,a\right)=T$$

(take $i_i:w\mapsto d_1$) and

$$V_{\mathcal{M}}\left(\exists x!\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right),w_1,a\right)=T$$

(take $i_i:w\mapsto d_2$). But

$$V_{\mathcal{M}}\left(F\left(\iota x\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right)\right)\wedge\neg\Box F\left(\iota x\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right)\right),@,a\right)=F.$$

Suppose that the latter is false. Then

$$V_{\mathcal{M}}\left(\neg\Box F\left(\iota x\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right)\right),@,a\right)=T,$$

which is the case if and only if it is not the case that

$$V_{\mathcal{M}}\left(\iota x\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right),w',a\right)\in I\left(F,w'\right)$$

for some $w'\in W$. Since there there are two worlds, there are two cases to be considered. First, note that $V_{\mathcal{M}}\left(\iota x\left(F\left(x\right)\wedge\neg\Box F\left(x\right)\right),@,a\right)=i_1\left(@\right)=d_1$. This is partly due to the fact that

$V_{\mathcal{M}}(F(x), @, a[i_1/x]) = T$, which can only be the case if $i_1(@) \in I(F, @)$ or, equivalently, $V_{\mathcal{M}}(\iota x(F(x) \wedge \neg\Box F(x)), @, a) \in I(F, @)$. Second, the case of w_1 is completely analogous. So, contrary to our assumption,

$$V_{\mathcal{M}}(F(\iota x(F(x) \wedge \neg\Box F(x))) \wedge \neg\Box F(\iota x(F(x) \wedge \neg\Box F(x))), @, a) = F$$

is true. ∎

Another possibility is to commute the box operator and the existential quantifier in the previous self-predication principle. This is a good move, since the resulting self-predication principle *is* provable in $\mathbf{C_T}$ *and* it holds for all well-formed formulas.

Theorem 4 $\vdash_{\mathbf{C_T}} \exists x\Box!\phi(x) \rightarrow \phi(\iota x\phi(x))$ *for all* ϕ □

Proof Assume that $\exists x\Box!\phi(x)$. Suppose that $\Box!\phi(t)$ for some term t, i.e. suppose that $\Box(\phi(t) \wedge \forall y(\phi(y) \rightarrow y = t))$. By $=$**I**, $\Box$-**T** and $\wedge$**I** one can infer from the previous that

$$\Box(\phi(t) \wedge \forall y(\phi(y) \rightarrow y = t)) \wedge \Box t = t.$$

It is a theorem of $\mathbf{C_T}$ that $(\Box\psi \wedge \Box\varphi) \rightarrow \Box(\psi \wedge \varphi)$. Consequently, one may derive that $\Box(\phi(t) \wedge \forall y(\phi(y) \rightarrow y = t) \wedge t = t)$. Since it is a theorem of $\mathbf{C_T}$ that $\Box\psi(t) \rightarrow \Box\exists x\psi(x)$, one may infer that $\Box\exists x(\phi(x) \wedge \forall y(\phi(y) \rightarrow y = x) \wedge x = t)$. Application of $\vee$**I** delivers one then the definiens of $\Box t = \iota x\phi(x)$ and, therefore, also the definiendum. One may use $\Box =$**E** to infer from $\Box t = \iota x\phi(x)$ and $\phi(t)$ that $\phi(\iota x\phi(x))$. Finally, it follows by $\exists$**E** from $\exists x\Box!\phi(x)$ that $\phi(\iota x\phi(x))$. ∎

Summing up, in this section I have discussed four description, all of which fall for obvious reasons in the category of what I have dubbed "self-predication" principles. The principles were in order of presentation:

$$\exists x!\phi(x) \rightarrow \phi(\iota x\phi(x)) \quad \text{for all non-modal } \phi \tag{1}$$

$$\exists x!\phi(x) \rightarrow \phi(\iota x\phi(x)) \quad \text{for all } \phi \tag{2}$$

$$\Box\exists x!\phi(x) \rightarrow \phi(\iota x\phi(x)) \quad \text{for all } \phi \tag{3}$$

$$\exists x\Box!\phi(x) \rightarrow \phi(\iota x\phi(x)) \quad \text{for all } \phi \tag{4}$$

The first and last are provable in $\mathbf{C_T}$, whereas the second and third are not. The above description principles will play a big role in the discussion to come.

In this section I have covered Carnapian modal logic. In the next section, I will take a dive into the subject of Carnapian modal arithmetic.

2 Carnapian Modal arithmetic

The subject of this section is the combination of the earlier introduced $\mathbf{C_T}$ on the one hand and arithmetic on the other hand. I will first look at the extension of $\mathbf{C_T}$ with the principles of elementary arithmetic, **Q**. Then I will turn to a further extension, namely the theory which results by adding the principles of induction. Call it **CPA**. The main part of this section is devoted to a proof of the statement that there is a collapse of modality into truth in **CPA**. I will analyze the result and provide a remedy.

The language of arithmetic, $\mathcal{L}^*$, consists of the individual constant $\mathbf{0}$, the one-place function $\mathbf{s}$, the two-place functions $+$, $\cdot$ and the two-place relation $<$. The standard interpretation of the language assigns the number zero to $\mathbf{0}$, the successor function to $\mathbf{s}$, the addition function to $+$, the multiplication function to $\cdot$, and the less-than relation to $<$. The theory of elementary arithmetic, $\mathbf{Q}$, consists of the axioms which describe the deductive behaviour of the forementioned individual, function and relation terms.

If one conjoins Carnapian logic with elementary arithmetic one gets Carnapian modal arithmetic, $\mathbf{CQ}$. The language of $\mathbf{CQ}$ contains $\mathcal{L}^* \cup \{\Box, \iota\}$. A standard model of $\mathcal{L}^* \cup \{\Box, \iota\}$ is a quintuple $\langle W, @, D, d^*, I\rangle$ as before, except for the following: $D = \mathbb{N}$ (with $\mathbb{N}$ the set of natural numbers), $d^* = n^*$ for some $n^* \in \mathbb{N}$, $I(\mathbf{0}) = 0$ (with 0 the number zero), $I(\mathbf{s}) = s$ (with s the successor function), $I(+) = +$ (with $+$ the addition function), $I(\times) = \times$ (with $\times$ the multiplication function) and with $I(<) =<$ (with $<$ the smaller than relation).

One could extend $\mathbf{CQ}$ by adding one of the principles of induction. The following are respectively the weak and the strong principle of induction.[4]

$$(\phi(\mathbf{0}) \wedge \forall x (\phi(x) \to \phi(\mathbf{s}(x)))) \to \forall x \phi(x) \tag{5a}$$

$$\forall x (\forall y (y < x \to \phi(y)) \to \phi(x)) \to \forall x \phi(x) \tag{5b}$$

When $\phi \in \mathcal{L}^*$ and when $\mathbf{Q}$ is the background theory, then the so-called "weak" principle of induction is deductively stronger than the so-called "strong" principle of deduction, since the weak principle of induction implies the strong principle of induction and since the strong principle of induction implies the weak principle of induction only if one assumes $\forall x (x = \mathbf{0} \vee \exists y (x = \mathbf{s}y))$, which is by the way also a consequence of the principle of weak induction. There are multiple textbook proofs of these deductive relations.[5] The reason for calling the deductively weaker principle the "strong" principle lies in the fact that the antecedent of the principle is stronger than is the antecedent of the deductively stronger principle. This can easily be seen from the fact that, given $\forall x (x = \mathbf{0} \vee \exists y (x = \mathbf{s}y))$, the antecedent of (5b) is equivalent to:

$$\phi(\mathbf{0}) \wedge \forall x (\forall y (y < \mathbf{s}(x) \to \phi(y)) \to \phi(\mathbf{s}(x)))$$

The second conjunct says in effect that one can deduce $\phi(\mathbf{s}(x))$ if one has $\phi(\mathbf{0})$, $\phi(\mathbf{s0})$, $\ldots$, $\phi(x)$. According to the second conjunct of the antecedent of (5a) it suffices to have $\phi(x)$. When $\phi \in \mathcal{L}^* \cup \{\Box\}$ and when $\mathbf{CQ}$ is the background theory, one must take care not to use the classical substitution principle in the proof, but it can be done.

I should note that the principle of strong induction (5b) is provably equivalent to the so-called "least number principle":

$$\exists x \phi(x) \to \exists x (\phi(x) \wedge \forall y (y < x \to \neg\phi(x))) \tag{6}$$

This principle will play an important in the next section.

It is usual to refer to $\mathbf{Q} \cup \{(5a)\}$ as "Peano Arithmetic" ($\mathbf{PA}$) and therefore it is appropriate to call $\mathbf{CQ} \cup \{(5a)\}$ "Carnapian Peano Arithmetic" ($\mathbf{CPA}$). I will now prove a lemma which will allow me to deduce that $\mathbf{CPA}$ collapses into $\mathbf{PA}$.

Lemma 1 $\vdash_{\mathbf{CPA}} \forall x (\Box x = \mathbf{0} \vee \exists y \Box (x = \mathbf{s}(y)))$ □

Proof First, one needs to prove that:

$$\Box \mathbf{0} = \mathbf{0} \vee \exists y \Box (\mathbf{s}(y) = \mathbf{0}) \tag{7}$$

But this is easy, since it follows from $=$**I** and $\Box$-**T**.

[4] The displayed formulas should be prefixed by as many universal quantifiers as there are free variables in ϕ, but for sake of brevity, I will suppress them.

[5] See e.g. Boolos et al. (2003, p. 212-214) for the relevant proofs.

Second, one needs to establish that:

$$\forall x \left(\left(\Box x = \mathbf{0} \vee \exists y \Box \left(\mathbf{s}\left(y\right) = x \right) \right) \rightarrow \left(\Box \mathbf{s}\left(x\right) = \mathbf{0} \vee \exists y \Box \left(\mathbf{s}\left(y\right) = \mathbf{s}\left(x\right) \right) \right) \right) \tag{8}$$

Assume that:

$$\Box t = \mathbf{0} \vee \exists y \Box \left(\mathbf{s}\left(y\right) = t \right)$$

We have to consider two cases. First, consider the case in which:

$$\Box t = \mathbf{0}$$

It is an instance of $=\mathbf{E}$ that $\forall x \forall y \left(x = y \rightarrow \mathbf{s}\left(x\right) = \mathbf{s}\left(y\right) \right)$ and, consequently, it is by successive application of $\forall\mathbf{E}$ and $\Box$-$\mathbf{T}$ it is also a theorem that $\Box t = \mathbf{0} \rightarrow \Box \mathbf{s}\left(t\right) = \mathbf{s}\left(\mathbf{0}\right)$. It therefore follows from $\Box t = \mathbf{0}$ that $\Box \mathbf{s}\left(t\right) = \mathbf{s}\left(\mathbf{0}\right)$. Existential generalization allows one then to conclude that $\exists y \left(\Box \mathbf{s}\left(t\right) = \mathbf{s}\left(y\right) \right)$. Second, consider the case in which:

$$\exists y \Box \left(\mathbf{s}\left(y\right) = t \right)$$

Suppose that there is some term t' such that $\Box \left(\mathbf{s}\left(t'\right) = t \right)$. From this one can derive by reasoning analogous to the above reasoning that $\Box \mathbf{s}\left(\mathbf{s}\left(t'\right)\right) = \mathbf{s}\left(t\right)$. One can apply existential generalization once more, the result being $\exists y \Box \left(\mathbf{s}\left(y\right) = \mathbf{s}\left(t\right) \right)$. There is no mentioning of t' anymore, so the latter existentially quantified formula follows from the former existentially quantified formula. So both cases lead one to conclude that $\exists y \Box \left(\mathbf{s}\left(y\right) = \mathbf{s}\left(t\right) \right)$. Hence, $\Box \mathbf{s}\left(t\right) = \mathbf{0} \vee \exists y \Box \left(\mathbf{s}\left(y\right) = \mathbf{s}\left(t\right) \right)$. This conclusion was reached for some arbitrary t. So one may apply universal generalization. This concludes the proof of (8).

The lemma follows then by modus ponens from (5a) and the conjunction of (7) and (8). ■

Corollary 1 $\vdash_{\mathbf{CPA}} \forall x \forall y \left(x = y \rightarrow \Box x = y \right)$ □

Proof First, one needs to prove the induction base:

$$\forall y \left(\mathbf{0} = y \rightarrow \Box \mathbf{0} = y \right) \tag{9}$$

So assume that $\mathbf{0} = t$ for some term t. Since it is an axiom of $\mathbf{Q}$ that $\neg \exists y \left(\mathbf{s}\left(y\right) = \mathbf{0} \right)$ and since it is by $=\mathbf{E}$ a consequence of the axiom and $\mathbf{0} = t$ that $\neg \exists y \left(\mathbf{s}\left(y\right) = t \right)$, it follows by the contraposition of $\Box\mathbf{E}$ that $\neg \exists y \Box \left(\mathbf{s}\left(y\right) = t \right)$. On the basis of lemma 1 one may conclude that $\Box t = \mathbf{0}$. It is a tautological consequence that $\mathbf{0} = t \rightarrow \Box t = \mathbf{0}$. Finally, universal generalization delivers us (9).

Second, one needs to prove the induction step:

$$\forall x \left(\forall y \left(x = y \rightarrow \Box x = y \right) \rightarrow \forall y \left(\mathbf{s}\left(x\right) = y \rightarrow \Box \mathbf{s}\left(x\right) = y \right) \right) \tag{10}$$

Assume that $\forall y \left(t = y \rightarrow \Box t = y \right) \wedge \mathbf{s}\left(t\right) = t'$, for some terms t and t'. From the conjunct $\mathbf{s}\left(t\right) = t'$ and from the $\mathbf{Q}$-axiom $\neg \exists y \left(\mathbf{s}\left(y\right) = \mathbf{0} \right)$, it follows by $=\mathbf{E}$ that $t' \neq \mathbf{0}$. Therefore it holds that $\neg \Box t' = \mathbf{0}$, from which by lemma 1 one can derive that $\exists y \Box \left(\mathbf{s}\left(y\right) = t' \right)$. Now suppose that $\Box \left(\mathbf{s}\left(t''\right) = t' \right)$, for some term t''. It is a consequence of the second conjunct, $\Box\mathbf{E}$ and $\mathbf{s}\left(t\right) = t'$ that $\mathbf{s}\left(t\right) = \mathbf{s}\left(t''\right)$. The $\mathbf{Q}$-axiom $\forall x \forall y \left(\mathbf{s}\left(x\right) = \mathbf{s}\left(y\right) \rightarrow x = y \right)$ guarantees then that $t = t''$. From here it follows by $\forall y \left(t = y \rightarrow \Box t = y \right)$ that $\Box t = t'''$. This allows one to use $(\Box =\mathbf{E})$ in order to derive $\Box \mathbf{s}\left(t\right) = t'$. Finally, (10) follows from $\rightarrow\mathbf{I}$, exportation and $\forall\mathbf{I}$.

The corollary follows from (9) and (10) by (5a). ■

Corollary 1 has tremendous importance, since it effectively reinstates the classical substitution principle in power. A very disturbing consequence of this is the collapse of **CPA** into **PA**.

Theorem 5 (Føllesdal) $\forall x \forall y (x = y \to \Box x = y) \vdash_{\mathbf{C_T}} (\phi \wedge \Box (z \neq c^*))$
$\to \Box\phi$ ◻

Proof Föllesdal (2004, p. 70) ■

Corollary 2 $\vdash_{\mathbf{CPA}} \phi \to \Box\phi$ ◻

Proof Take $\mathbf{s}(c^*)$ for z. It is a theorem of **CQ** that $\forall x \Box (x \neq \mathbf{s}(x))$ ■

Is this result relevant? Most philosophers are convinced of the necessity of arithmetical truths, so they will prima facie not be bothered by the above result. Recall however that the language *contained* $\mathcal{L}^* \cup \{\Box, \iota\}$ but is *not identical* to it. To be more precise, the language also contains predicates and relations other than $<$. Among these predicates and relations there might be predicates contingently satisfied by numbers. But more importantly, one should not automatically assume that the box operator should be read as a necessity operator. For instance, in the next section we will look at a system which contains **S4**. The **S4**-laws can be seen as epistemic principles. Since **S4** contains **T**, the collapse result holds also for $\mathbf{C_{S4}PA}$. And it is far more controversial to claim that it is all right for epistemic modalities to collapse into truth.

So Carnapian Peano Arithmetic is in a bad shape. Now I turn to a diagnosis. First, I will prove that the standard weak principle of induction is invalid. Then I will propose an alternative weak principle of induction. This alternative principle will be proved to be valid. Next, I will show that there is no collapse of modality into truth if one extends **CQ** with the alternative principle of induction. Finally, as a bonus I will show that one can retain the standard least number principle.

As stated above, the standard weak principle of induction is invalid.

Theorem 6 *There is a standard model of* **CQ** *in which (5a) is not true.* ◻

Proof Consider any standard model of **CQ** with W containing at least two possible worlds and with R a total relation on W. Let ϕ be the following well-formed formula:

$$\Box x = \mathbf{0} \vee \exists y \Box x = \mathbf{s}(y)$$

One can then show that:

1. $V_{\mathcal{M}} (\phi(\mathbf{0}) \wedge \forall x (\phi(x) \to \phi(\mathbf{s}(x))), a) = T$;
2. $V_{\mathcal{M}} (\forall x \phi(x), a) = F$.

The first statement is true if only if both conjuncts are true.

Consider the first conjunct. $V_{\mathcal{M}} (\Box \mathbf{0} = \mathbf{0} \vee \exists y \Box \mathbf{0} = \mathbf{s}(y), a) = T$. It suffices that one of the disjuncts is true. Now $V_{\mathcal{M}} (\Box \mathbf{0} = \mathbf{0}, a) = T$ if and only if $V_{\mathcal{M}} (\mathbf{0} = \mathbf{0}, w, a) = T$ in all words $w \in W$, which at its turn is true if and only if $I(\mathbf{0}) = I(\mathbf{0})$. The latter holds in any model, so one of the disjuncts of the first conjunct is true and, ergo, the first conjunct is true.

Consider the second conjunct:

$$V_{\mathcal{M}} (\forall x ((\Box x = \mathbf{0} \vee \exists y \Box x = \mathbf{s}(y))) \to (\Box \mathbf{s}(x) = \mathbf{0} \vee \exists y \Box \mathbf{s}(x) = \mathbf{s}(y)), a) = T.$$

The above is true if and only if for all i, if $i(w) = 0$ for all $w \in W$ or there is a i' such that

$$i(w) = V(\mathbf{s}(y), w, a[i'/y])$$

for all $w \in W$, then $V_{\mathcal{M}} (\mathbf{s}(x), w, a[i/x]) = 0$ for all $w \in W$ or there is a i' such that $V_{\mathcal{M}} (\mathbf{s}(x), w, a[i/x]) = V(\mathbf{s}(y), w, a[i'/y])$ for all $w \in W$. Assume that the antecedent holds.

Furthermore, suppose that the first disjunct holds. Then $V_{\mathcal{M}}(\mathbf{s}(x), w, a[i/x]) = 1$ for all $w \in W$. Consequently, there is a i' such that

$$V_{\mathcal{M}}(\mathbf{s}(x), w, a[i/x]) = V(\mathbf{s}(y), w, a[i'/y])$$

for all $w \in W$. Indeed, take $i' : w \mapsto 0$. If one disjunct of the consequent is true, then the consequent itself is also true. Next, suppose that the second disjunct holds. Then there is a i' such that $i(w) = V(\mathbf{s}(y), w, a[i'/y])$ for all $w \in W$. Hence there is a i' such that $V_{\mathcal{M}}(\mathbf{s}(x), w, a[i/x]) = V(\mathbf{s}(\mathbf{s}(y)), w, a[i'/y])$ for all $w \in W$. Now if $V_{\mathcal{M}}(\mathbf{s}(y), w, a[i'/y]) = n$, let $i'' : w \mapsto n$. Then there is a i'' such that $V_{\mathcal{M}}(\mathbf{s}(x), w, a[i/x]) = V(\mathbf{s}(y), w, a[i'/y])$ for all $w \in W$. Again, if one disjunct of the consequent is true, then the consequent itself is also true. Moreover, if the consequent follows from either of the disjuncts of the antecedent, then it follows from the disjunction too. Thus the second conjunct is also true.

The second statement is true if and only if there is an i such that $i(w') \neq 0$ for some world $w' \in W$ and for all i' there is some world w'' such that $i(w'') \neq V(\mathbf{s}(y), w'', a[i'/y])$. Let $i(w') = 1$ and $i(w'') = 0$. Clearly, $i(w') = 1 \neq 0$. Moreover, $i(w'') = 0 \neq V(\mathbf{s}(y), w'', a[i'/y])$, whatever i' may assign to y in w''.

This concludes the proof. ∎

The interesting thing about the above proof is not only that it shows that the standard weak principle of induction is invalid, but also that that induction principle fails precisely in the case in which it was used in the proof of theorem 1, which was essential in the larger proof of the collapse result.

If the standard weak principle of induction is invalid, then one should wonder if there is a valid induction principle which can replace it. Recall that (5a) was supposed to hold for all $\phi \in \mathcal{L}$. Therefore the substitution principle ($\Box = \mathbf{E}$) is relevant. This teaches us that the antecedent of (5a) is equivalent to the following:

$$\forall x(\Box x = \mathbf{0} \to \phi(x)) \wedge \forall x(\forall y(\Box y = x \to \phi(y)) \to \forall y(\Box y = \mathbf{s}(x) \to \phi(y))) \quad (11)$$

A similar point can be made about the antecedent of (5b), which is equivalent to:

$$\forall x (\forall y (y < x \to \phi(y)) \to \forall y (\Box y = x \to \phi(y))) \quad (12)$$

Now the idea is to change the antecedent of the weak induction principle as follows:

$$\forall x(x = \mathbf{0} \to \phi(x)) \wedge \forall x(\forall y(y = x \to \phi(y)) \to \forall y(y = \mathbf{s}x \to \phi(y))) \quad (13)$$

In a similar vein the antecedent of (5b) could then be rewritten as follows.

$$\forall x (\forall y (y < x \to \phi(y)) \to \forall y (y = x \to \phi(y))) \quad (14)$$

The latter is provably equivalent to:

$$\exists x (\exists y (x = y \wedge \phi(y)) \wedge \forall y (y < x \to \neg\phi(y))) \quad (15)$$

Whether this is a good idea depends, first, on whether (13) $\to \forall x \phi(x))$ (call it **W-INDa**) is a valid principle, second, whether one manages to block the proof of the collapse result, and, third, whether modalities do not collapse in $\mathbf{CQ} \cup \{\mathbf{W-INDa}\}$ (call this theory **CMA**). I can be short about the second requirement. The proof of corollary 2 is blocked, since one cannot prove that $\forall x (x = \mathbf{0} \to \Box x = \mathbf{0})$, as is needed to prove theorem 1. The other requirements are also met, as is proved below.

Theorem 7 ***W-INDa*** *is valid.* □

Proof The proof is by induction on the complexity of ϕ. I will restrict myself to the case in which $\phi = \Box\psi$. One can prove that the theorem holds in this case by reductio ad absurdum. So assume that: $V_{\mathcal{M}}\left((13), @a\right) = T$, and $V_{\mathcal{M}}\left(\forall x \phi\left(x\right), @, a\right) = F$. The second statement is true if and only if $V_{\mathcal{M}}\left(\psi\left(x\right), w', a[i'/x]\right) = F$ for some i' and some $w' \in W$. By definition $V_{\mathcal{M}}\left(x, @, a[i'/x]\right) \in \mathbb{N}$. Suppose that $V_{\mathcal{M}}\left(x, @, a[i'/x]\right) = n$ for some $n \in \mathbb{N}$. Moreover, the first statement makes it true that for all i'', if $V_{\mathcal{M}}\left(x, @, a[i''/x]\right) = n$, then $V\left(\phi\left(x\right), w'', a[i''/x]\right)$ for all $w'' \in W$. So it is also true that if

$$V_{\mathcal{M}}\left(x, @, a[i'/x]\right) = n,$$

then $V\left(\phi\left(x\right), w', a[i'/x]\right) = T$. The consequent is false. Hence, $i'\left(@\right) \neq n$, contrary to the assumption. But n was arbitrary. Consequently, it is not the case that $V_{\mathcal{M}}\left(x, @, a[i'/x]\right) \in \mathbb{N}$. Contradiction. ■

Theorem 8 *There is a standard model of* **CQ** *in which some well-formed ϕ is true but $\Box\phi$ is not.* □

Proof Consider the following model: $D = \mathbb{N}$; $W = \{@, w\}$; R is a total relation on W; $I\left(P, @\right) = \{2\}$; $I\left(P, @\right) = \{3\}$; the standard assignments to the arithmetical vocabulary (see above). The following two claims hold then: $V_{\mathcal{M}}\left(P\left(\mathbf{s}\left(\mathbf{s}\left(\mathbf{0}\right)\right)\right), @, a\right) = T$ and, hence,

$$V_{\mathcal{M}}\left(P\left(\mathbf{s}\left(\mathbf{s}\left(\mathbf{0}\right)\right)\right)\right) = T;$$

$V_{\mathcal{M}}\left(P\left(\mathbf{s}\left(\mathbf{s}\left(\mathbf{0}\right)\right)\right), w, a\right) = F$ and, therefore, $V_{\mathcal{M}}\left(\Box P\left(\mathbf{s}\left(\mathbf{s}\left(\mathbf{0}\right)\right)\right), @, a\right) = F$, and, hence,

$$V_{\mathcal{M}}\left(\Box P\left(\mathbf{s}\left(\mathbf{s}\left(\mathbf{0}\right)\right)\right)\right) = F.$$

It was argued that it is a good move to replace the antecedent of the standard weak principle of induction by (13). One could wonder at this point whether one should also replace the antecedent of the standard strong principle of induction by (14) or, equivalently, replace the consequent of the standard least number principle by (15). The answer is negative. Once one has $\mathbf{W} - \mathbf{INDa}$, one also has the standard strong principle of induction.

Theorem 9 $\mathbf{W} - \mathbf{INDa} \vdash_{\mathbf{C_T}} (6)$ □

Proof Take the proof of $(5a) \vdash (6)$ in Boolos et al. (2003, p. 213-214). There it is assumed that $\neg\exists x\left(\phi\left(x\right) \wedge \forall y\left(y < x \rightarrow \neg\phi\left(y\right)\right)\right)$. Then one can prove that $\forall x \neg\exists y\left(y < x \wedge \phi\left(y\right)\right)$. The proof is by induction. At this point one should replace (5a) with $(13) \rightarrow \forall x \phi\left(x\right)$. It is easy to check that

$$\forall x\left(x = \mathbf{0} \rightarrow \neg\exists y\left(y < x \wedge \phi\left(y\right)\right)\right)$$

is true for the same reason that $\neg\exists y\left(y < \mathbf{0} \wedge \phi\left(y\right)\right)$ is true. Next, one needs to show that

$$\forall x(\forall z(z = x \rightarrow \neg\exists y(y < z \wedge \phi(y))) \\ \rightarrow \forall z(z = \mathbf{s}(x) \rightarrow \neg\exists y(y < z \wedge \phi(y))))$$

The proof is by reductio ad absurdum. So suppose that for some term t,

$$\forall z\left(z = t \rightarrow \neg\exists y\left(y < z \wedge \phi\left(y\right)\right)\right),$$

for some term t', $t' = \mathbf{s}(t)$, and for some term t'', $t'' < t'$ and $\phi(t'')$. Then $t'' < \mathbf{s}(t)$. Hence, by an axiom of arithmetic, $t'' = t$ or $t'' < t$. Suppose the latter. It follows by $\exists\mathbf{I}$ that $\exists y\,(y < t \wedge \phi(y))$. The first of the reductio assumptions and $=\mathbf{E}$ imply that $\neg\exists y\,(y < t \wedge \phi(y))$. So one must drop that option. Suppose then that $t'' = t \wedge \phi(t'')$. It follows from the first conjunct that $\neg\exists y\,(y < t'' \wedge \phi(y))$. One may then infer on the basis of $\wedge\mathbf{I}$ and $\exists\mathbf{I}$ that

$$\exists x\,(\phi(x) \wedge \forall y\,(y < x \rightarrow \neg\phi(y))),$$

contradicting the very first assumption of the proof. The remainder of the proof is exactly as in Boolos et al. (2003, p. 214). ■

This concludes the discussion of Carnapian modal arithmetic. Next on the agenda is Carnapian epistemic arithmetic.

3 Carnapian Epistemic Logic and Arithmetic

Horsten (2005) described and used Carnapian Epistemic Arithmetic (henceforth **CEA**) in an interesting and fruitful manner. I will very briefly write about the language of **CEA**, its informal and informal interpretation, the theory and its consistency.

In his paper Horsten starts with second-order languages, whereas in this paper I started with first-order languages. In order not to complicate things, I will stick to the first-order fragment of his theory. Moreover, in this section I will ignore descriptions. The informal interpretation of the box operator is as follows:

> The operator $\Box$ will be interpreted as "It can be established in the actual world by person K that", "It is verifiable in the actual world by person K that", or "It can be shown in the actual world by person K that". Here the person K, our nondescript epistemic agent, will be kept fixed throughout. Horsten (2005, p. 233-234)

About the formal interpretation he says the following:

> A model $\mathcal{M}$ for a Carnapian language $\mathcal{L}_C$ is determined by a universe of objects $\mathcal{D}$, a collection of presentations (of objects, sets, relations) taken from a language $\mathcal{L}$, a collection of sentences S of this language $\mathcal{L}$, and assignments to expressions of $\mathcal{L}_C$. The elements of S make up the extension of $\Box$ in $\mathcal{M}$: the elements of S are the sentences the truth of which our epistemic agent K is able to establish. Truth in a model, validity, consequence are then defined on the basis of such a notion of models for Carnapian languages. Horsten (2005, p. 234)

Horsten freely admits that there is more formal work to be done here. But the basic idea is more or less clear. If $\exists x \Box F(x)$ is true, then there is a presentation p_o of an object o such that $F(p_0)$ is in S. A presentation p_o should be presumably be thought of as a closed term of some sort.

The common feature of the formal interpretation intended by Horsten and the Carnap-Kremer interpretation consists in what Kripke dubbed "Carnapian double think". In the Carnap-Kremer interpretation variables range over individual concepts, i.e. functions from the domain to possible worlds. But if a formula does not contain box operators, then only the values of the individual concepts at a given world, viz. the objects, count. In the Horsten interpretation variables range over presentations of objects when they are in the scope of a box operator and they range over the objects themselves when they are embedded in non-epistemic contexts.

Carnapian epistemic logic is just like $\mathbf{C_T}$, except that $\Box$-$\mathbf{T}$ is replaced by:

$\Box$**I-S4** If $(\Box)\,\Gamma \vdash \phi$, then $(\Box)\,\Gamma \vdash \Box\phi$.

Call the theory which is exactly like $\mathbf{C_T}$, except that $\Box$**I-T** is replaced by $\Box$**I-S4**, $\mathbf{C_{S4}}$. $\mathbf{C_{S4}}$ encompasses $\mathbf{C_T}$. Horsten proved the consistency of $\mathbf{C_{S4}}$ by making use of the eraser translation. Of course, the Carnapian models used in the previous sections of this paper will do just fine for a soundness proof and thus for a consistency proof. The only thing which needs to be taken care for is that the relation R should not merely be a reflexive two-place relation but also a transitive reflexive two-place relation. Obviously, since there is not a full-detailed formal interpretation of the lines sketched by Horsten, there cannot be a soudness nor a completeness proof with the models intended by him.

As for informal soundness, there are arguments to be found in the literature that the sentential part of $\mathbf{C_{S4}}$ (call it **S4**) it is indeed the correct logic for something like demonstrability.[6] It should be noted that it is desirable to restrict $=\mathbf{E}$ to non-modal formulas, so as to avoid the provability of $\forall x \forall y\,(x = y \rightarrow \Box x = y)$, since the latter is informally incorrect given the intended reading and given its instantiability with description terms.

The combination of $\mathbf{C_{S4}}$ with **PA** is called $\mathbf{C_{S4}PA}$. One might ask oneself the question whether the standard weak principle of induction is valid in the Horsten interpretation. This is a pressing question, since the reasoning developed in section 2 can be repeated here to demonstrate that the standard weak principle of arithmetic leads to a collapse of the notion of provability or verifiability into truth. Interestingly, Horsten makes the following comment about the relevant principle of induction:

> It should be noted that the full second-order principle of mathematical induction is not intuitively valid in the Carnapian setting. The antecedent of the induction axiom is satisfied by a property Θ if Θ holds of the numerals 0, $S0$, $SS0$, ...But for the consequent to be valid Θ must hold not only for these standard numerals, but for *all* presentations of natural numbers. In response to this difficulty, one may want to restrict universal quantifier in the consequent of the induction axiom to the *standard Peano numerals*. (Horsten, 2005, p. 256)

This is a problem which is analogous to the one we encountered before: the antecedent of the induction principle says that $\phi\,(x)$ holds of all the natural numbers, but the consequent says that $\phi\,(x)$ holds of all the individual concepts. Horsten could have opted for a solution analogous to the one given in the previous section. Indeed, under his interpretation (13) says that ϕ is true of all number denoting terms, which are of course all the terms, not only the standard numerals or the terms which provably codenote with the standard numerals. But instead he opted for a different solution. Rather than making the range of the x in the $\phi\,(x)$ of the antecedent of the induction principle less restrictive, he suggests to restricting the range of the x in the $\phi\,(x)$ more restrictive. In a first-order setting one could achieve this by replacing the consequent $\forall x \phi\,(x)$ by $\forall x \exists y\,(x = y \wedge \phi\,(y))$.[7] Let us call the resulting alternative principle of weak induction **W-INDc**. Note that **W-INDc** also allows for a proof of (6). This is good, since Horsten argues that (6) should be valid under his interpretation.[8] Not only do **W-INDa** and **W-INDc** seem to be equally valid and equally good for proving the least number principle, they also both block the collapse argument expounded in section 2. So I lack a reason to prefer **W-INDa** to **W-INDc**.

In the next section we will see what happens if one adds self-predication principles to Carnapian Epistemic Arithmetic.

[6] See Burgess (1999).
[7] This was pointed out to me in private communication with Horsten.
[8] See Horsten (2008).

4 Descriptions in Carnapian Epistemic Arithmetic

In subsection 1.3 I discussed self-predication principles. One might ask oneself if one of the previously discussed self-predication principles are valid under the Horsten interpretation, if there are any. If the descriptive condition is non-epistemic, then what the restricted self-prediction principle (2) says is that, if there is some object such that the descriptive condition ϕ is uniquely satisfied by it, then the descriptive condition is also satisfied by the object denoted by the description term. This seems to be all right. If the descriptive condition is epistemic, then what the unrestricted self-prediction principle (2) says is that, if there is some closed term t such that the descriptive condition ϕ is satisfied by t and such that all terms t' which satisfy the descriptive condition are codenoting with t, then the descriptive condition ϕ is also satisfied by the description term $\iota x\phi$ itself. Of course, t need not have been $\iota x\phi$, so there does not seem to be a necessary connection between the antecedent and the consequent. So perhaps the unrestricted self-predication principle is not valid either in Carnapian Epistemic Arithmetic. But pending a formal Horsten interpretation it is perhaps safer to take a look at some of the consequences of the self-predication principles. If there are some consequences, e.g. collapse results, which show that the principles must be informally unsound, then we have an additional reason to drop them.

One interesting set of consequences from the above point of view can be derived using the following lemma:

Lemma 2

$$\vdash_{\mathbf{CQ}} \exists x(\phi(x) \wedge \forall y(y < x \to \neg\phi(y))) \\ \to \exists x!(\phi(x) \wedge \forall y(y < x \to \neg\phi(y))) \quad \square$$

Proof Left as an excercise for the reader. ■

The self-predication principle (1) which is restricted to non-modal/non-epistemic well-formed formulas has the following consequence:

Theorem 10 $\vdash_{\mathbf{C_T Q}\cup\{(1),(6)\}} \Box\exists x\phi\,(x) \to \exists x\Box\phi\,(x)$ *for all non-modal* ϕ □

Proof Assume first that $\exists x\phi\,(x)$. The least-number principle (6) allows one to infer

$$\exists x\,(\phi\,(x) \wedge \forall y\,(y < x \to \neg\phi\,(y)))$$

from $\exists x\phi\,(x)$. Lemma 2 guarantees then that $\exists x!\,(\phi\,(x) \wedge \forall y\,(y < x \to \neg\phi\,(y)))$. Let us abbreviate

$$\phi\,(x) \wedge \forall y\,(y < x \to \neg\phi\,(y))$$

with $\psi\,(x)$. One may use (1) to prove that $\psi\,(\iota x\psi\,(x))$. Subsequently, one may put $\wedge$**E** to work to derive $\phi\,(\iota x\psi\,(x))$. By $\Box$**I-T** one can derive $\Box\phi\,(\iota x\psi\,(x))$ from $\Box\exists x\phi\,(x)$. All that leaves to be done then, is to apply $\exists$**I** to the latter, the result being a proof of $\exists\Box\phi\,(x)$ from the assumption that $\Box\exists\phi\,(x)$. ■

Analogous theorems involving the unrestricted self-predication principle (2) and one of the weaker unrestricted self-predication principles, (3), can be proved analogously.

Theorem 11 $\vdash_{\mathbf{C_T Q}\cup\{(2),(6)\}} \Box\exists x\phi\,(x) \to \exists x\Box\phi\,(x)$ *for all* ϕ □

Theorem 12 $\vdash_{\mathbf{C_{S4} Q}\cup\{(3),(6)\}} \Box\exists x\phi\,(x) \to \exists x\Box\phi\,(x)$ *for all* ϕ □

Obviously, no such theorem can be proved with (4) as the sole self-predication principle.

The formula featuring in the above theorems is sometimes called the converse Ghilardi formula. It is not valid under the Kremer interpretation. Shapiro pointed out its invalidity to Marti. He noted, more specifically, that if there is more than one object in the universe, then $\Box\exists x\,(x \neq c \wedge \Diamond\neg x \neq c)$ is true but $\exists x\Box\,(x \neq c \wedge \Diamond\neg x \neq c)$ is not. It is no coincidence that this counterexample has the same form as my counterexample to (3). However, as I stressed above, as long as there is no formal Horsten interpretation, we should look at other means of judging the soundness of the principle involved.

The proofs of the above theorems can be easily adapted to proofs for the Barcan formula:[9]

Theorem 13 $\vdash_{\mathbf{C_T Q}\cup\{(1),(6)\}} \Diamond\exists x\phi\,(x) \to \exists x\Diamond\phi\,(x)$ *for all non-modal* ϕ □

Theorem 14 $\vdash_{\mathbf{C_T Q}\cup\{(2),(6)\}} \Diamond\exists x\phi\,(x) \to \exists x\Diamond\phi\,(x)$ *for all* ϕ □

Theorem 15 $\vdash_{\mathbf{C_{S4} Q}\cup\{(3),(6)\}} \Diamond\exists x\phi\,(x) \to \exists x\Diamond\phi\,(x)$ *for all* ϕ □

This has the prima facie devastating consequence that *de re* verifiability collapses into *de dicto* verifiability, since $\exists x\Box\phi\,(x) \to \Diamond\exists x\phi\,(x)$ and $\Box\forall x\phi\,(x) \to \forall\Box x\phi\,(x)$ are theorems too.

Ackerman (1979) gave an argument which is in important respects similar to the the kernel of the proof of theorem 10, and she thought that the argument is invalid. I will expand her argument so as to make the relevancy of it more apparent. Suppose that someone believes that there are perfect numbers, i.e. numbers which are equal to the sum of their proper divisors. Then that someone believes that there is a least perfect number, which is of course unique, and therefore he or she believes that the least perfect number is the least perfect number. It follows then by existential generalization that there is a number of which he or she believes that it is the least perfect number. But, she added, that conclusion is wrong. It does not need to be the case that he or she believes of the number six, which happens to be the least perfect number, that it is the least perfect number. In her eyes the argument allowed an unacceptable transition from a *de dicto* statement to a *de re* statement. Her preferred solution was to restrict existential generalization to standard numerals. Ackerman's argument is interesting, but perhaps not very persuasive, since belief is not closed under logic, as is witnessed by the notorious inconsistency and incoherence of people's beliefs. Belief is just too liable to all sorts of intrinsic and extrinsic factors: failing memories, limited intelligence, lack of time, inadequate training, political correctness, and so on. In contrast, absolute provability is closed under first-order logic. Thus instead of following Ackerman's lead by sacrifizing the unrestricted principle of existential generalization, one could just as well or even better deny the closure of belief under modus ponens.

But whatever the merits of her argument are, it is under the Horsten interpretation of the variables and quantifiers very questionable to see $\exists x\Box\phi\,(x)$ as an expression of a *de re* epistemic (or in her case, doxastic) attitude towards numbers. The standard syntactic definition of a *de re* formula stipulates that a formula is *de re* if and only if there is a free variable occurring in the scope of a modal operator. Under this definition $\exists x\Box\phi\,(x)$ is *de re*. But the definition is correct only given the interpretation of variables as ranging over objects (even) when the variables occur in the scope of a modal operator. In the Horsten interpretation variables ranges not only over objects but also over closed terms denoting those objects, and when the variables occur in the scope of a modal operator the terms are the values assigned to the variables. Thus what $\exists x\Box\phi\,(x)$ says is that there is some closed, denoting term t such that it is absolutely provable that $\phi\,(t)$. Moreover, if one scrutinizes the proof of theorem 10, it should be apparent that what is actually proved is that there is a denoting closed *descriptive* term, $\iota x\psi\,(x)$, such that is absolutely provable that $\phi\,(\iota x\psi\,(x))$. It is central to the broadly Carnapian approach to epistemic

[9] I would like to thank Horsten for directing my attention to this fact.

logic that the coreferentiality of a descriptive term and some other denoting closed term does not in general entail the provability of the coreferentiality statement. To sum up, it is doubtful whether $\exists x \Box \phi(x)$ expresses *de re* absolute provability, since: (i) the value of x is not an object, but a term; and (ii) its value is not merely a term, but a descriptive term, which need not be epistemically transparant.

As for Ackerman's solution to the problem, namely to restrict the quanfifier principles to standard numerals rather than all terms, it should be noted that although it is indeed possible to restrict the quantifier principles to description-free terms, there are alternatives. For instance, one could refuse to manipulate the quantifier principles and adopt (4) as the sole self-predication principle when the well-formed formula involved is epistemic.

Theorem 11 has an interesting consequence:

Corollary 3 $\vdash_{\mathbf{C_T Q} \cup \{(2),(6)\}} \neg \Box \exists x (\phi(x) \wedge \neg \Box \phi(x))$ □

Proof Suppose that $\Box \exists x (\phi(x) \wedge \neg \Box \phi(x))$. By theorem 3 it follows that $\exists x \Box (\phi(x) \wedge \neg \Box \phi(x))$. By $\Box$**I-T**, one may conclude that $\Box(\psi \wedge \varphi) \vdash \Box \psi$ from $\psi \wedge \varphi \vdash \psi$, and the latter is easily proved from the rule of assumptions and $\wedge$**E**. So we also have $\exists x (\Box \phi(x) \wedge \Box \neg \Box \phi(x))$. Applying $\Box$**E** leads straight to the contradictory $\exists x (\Box \phi(x) \wedge \neg \Box \phi(x))$. ■

Again, similar corollaries can be derived given the other self-predication principles, if the background epistemic logic is strong enough. Corollary 3 has been proven by Horsten (2005, p. 238) with the difference that he did not prove theorem 11 but rather gave a direct proof of the above result. In the same paper he concluded from this that (2) is unsound.

For further discussion of this it is convenient to replace the unitary knowability operator $\Box$ by a complex operator $\Diamond K$, with $\Diamond$ the possibility operator and K the knowledge operator. Let us call the language with the unitary knowability operator $\mathcal{L}_{\Box}$ and the other language $\mathcal{L}_{\Diamond K}$. The **S4**-logic could then be replaced by a logic in which every occurence of $\Box$ in a **S4**-theorem is replaced by an occurrence of $\Diamond K$, plus the following purely epistemic principles:

FACT $K\phi \rightarrow \phi$

DIST $K(\phi \wedge \psi) \rightarrow (K\phi \wedge K\psi)$

It will be convenient to use the following abbreviatory definitions:

Definition 4 $\phi^*(x) =_{df} \phi(x) \wedge \neg K\phi(x)$ □

Definition 5 $\phi^{**}(x) =_{df} \phi^*(x) \wedge \neg \Diamond \phi^*(x) \wedge \forall y (y < x \rightarrow \neg \phi^*(y))$ □

In this logic the following lemma holds:

Lemma 3 $\Diamond K \forall x \neg \Diamond K \phi^*(x)$ □

Proof Suppose for a reductio that $\Diamond K \phi^*(t)$, for some arbitrary t. Use **DIST** to deduce $\Diamond(K\phi(t) \wedge K\neg K\phi(x))$. Then use **FACT** to deduce the contradictory $\Diamond(K\phi(t) \wedge \neg K\phi(x))$. Finally, use universal generalization and the fact that $\Diamond K$ is closed under theoremhood. ■

The above lemma can be put to good use in the proof of the following theorem:

Theorem 16 $\vdash_{(2),(6)} \neg \Diamond K \exists x (\phi(x) \wedge \neg K\phi(x))$ □

Proof Assume for a reductio that $\Diamond K \exists x \phi^* (x)$. Use the lemma and the distributivity of $\Diamond K$ over material implication to get

$$\Diamond K \exists x \left(\phi^* (x) \wedge \neg \Diamond K \phi^* (x)\right).$$

Next, use (6) and the same $\Diamond K$-rule to get $\Diamond K \exists x! \phi^{**} (x)$. Application of (2) yiels $\phi^{**} (\iota x \phi^{**} (x))$, which entails $\Diamond K \phi^* (\iota x \phi^{**} (x))$. Reasoning analogous to the reasoning in the proof of the lemma leads then to a contradiction. ■

In proof I make use of the distributivity of $\Diamond K$ over material implication. However, it is possible to give a counterexample to that modal-epistemic principle: (16) and (17) could both be true, whereas (18) cannot be.[10]

$$\Diamond K K (\mathbf{0} = \mathbf{0}) \tag{16}$$
$$\Diamond K \neg K (\mathbf{0} = \mathbf{0}) \tag{17}$$
$$\Diamond K \left(K (\mathbf{0} = \mathbf{0}) \wedge \neg K (\mathbf{0} = \mathbf{0})\right) \tag{18}$$

Note that it is crucial for both the counterexample and for the proof of the theorem that the complex operator $\Diamond K$ may be split up in the simple operators $\Diamond$ and K. So one line of response to the above argument against (2) is restricting it to formulas in which every K is preceded by a $\Diamond$.[11] It is supposed to be non ad hoc, since the move is also needed to prevent counterexamples to other principles of the epistemic logic, e.g. the distributivity of $\Diamond K$.[12]

If a collapse of *de re* into *de dicto* verifiability already worries certain people, then a collapse of verifiability into truth will worry many more. As far as I can see, there is no way to prove such a collapse on the basis of the self-predication principles alone. But it can be done, if the following rule also belongs to the system:

$\iota\mathbf{E}$ If $\Gamma \vdash \iota x \phi (x) \neq c^*$, then $\Gamma \vdash \exists x! \phi (x)$.

When taken as primitive rules the self-predication principles are in a way description introduction rules. Normally one should then also have description elimination rules. The above rule falls within that latter category. It is a valid rule within the Kremer interpretation. With the unrestricted self-predication principle and the above rule one can then prove the following theorem.

Theorem 17 *(2), $\iota\mathbf{E} \vdash (\phi \wedge \exists z \Box z \neq c^*) \rightarrow \Box \phi$ for all closed well-formed formulas ϕ* □

Proof Suppose that $\phi \wedge \exists z \Box z \neq c^*$, for any closed ϕ. By $=\mathbf{I}$ and $\wedge\mathbf{I}$ one may infer from $\Box t \neq c^*$ that $\Box t \neq c^* \wedge t = t$. By $\exists\mathbf{I}$, $\wedge\mathbf{I}$, and the fact that ϕ is closed, it follows that

$$\exists z \left(\phi \wedge \exists x \left(\Box z \neq c^* \wedge z = x\right)\right)$$

Ergo, the above follows by $\exists\mathbf{E}$ from $\phi \wedge \exists z \Box z \neq c^*$. Clearly, the latter entails that

$$\exists z! \left(\phi \wedge \exists x \left(\Box z \neq c^* \wedge z = x\right)\right)$$

For suppose not. Then there is some y such that it inter alia satisfies $y = x$ but also $y \neq z$. But since $z = x$ is also given, this is a violation of transitivity of identity. So the condition for

[10] See Horsten (2008). He also gave a more complicated counterexample to the principle in higher-order settings in his Horsten (2000, p. 50-51).

[11] This was suggested by Horsten (2008).

[12] See Horsten (2008).

deriving

$$\phi \wedge \exists x(\Box \iota z(\phi \wedge \exists x(\Box z \neq c^* \wedge z = x)) \neq c^* \\ \wedge\, \iota z(\phi \wedge \exists x(\Box z \neq c^* \wedge z = x)) = x) \quad \blacksquare$$

on the basis of (2) is fulfilled. From the fact that x is not free for quantification in

$$\Box \iota z\,(\phi \wedge \exists x\,(\Box z \neq c^* \wedge z = x)) \neq c^*$$

one can infer by $\wedge$**E** and $\exists$**E** that

$$\Box \iota z\,(\phi \wedge \exists x\,(\Box z \neq c^* \wedge z = x)) \neq c^*$$

The combination of ι**E**, $\Box$**I**, and the last displayed formula gives one:

$$\Box \exists z!\,(\phi \wedge \exists x\,(\Box z \neq c^* \wedge z = x))$$

Corollary 4 *(2), ι**E**,* $\mathbf{Q} \vdash \phi \to \Box\phi$ $\Box$

There are of course ways to avoid a collapse. One way would be to drop (2) and adopt either (3) or (4) for the cases in which the descriptive conditions are epistemic, or even adopt no other self-predication principle than (1). Adopting (4) has already been considered as a solution for avoiding a collapse of *de re* into *de dicto* verifiability. Another possibility is to tinker with the elimination rule for descriptions. In Horsten (2005, p. 238) the following description rule is considered as the only primitive rule for descriptions: from $\exists x!\phi\,(x)$, infer $\forall y\,(y = \iota x\phi\,(x) \leftrightarrow !\phi\,(y))$.[13] It is easy to see that the rule in question implies (2), but at the same time there is no collapse forthcoming from the combination of (2) with the derived elimination rule (if $\Gamma \vdash \exists x!\phi\,(x) \wedge \iota x\phi\,(x) = t$, then $\Gamma \vdash !\phi\,(t)$).

Summing up, adopting the unrestricted self-predication principle (2) leads to a collapse of *de re* into *de dicto* demonstrability or verifiability and, if one adopts certain elimination rules for descriptions, one even gets a collapse of demonstrability or verifiability into truth. While the first result need not be a bad result given the Horsten interpretation, the second result certainly is. Both kinds of collapse can be avoided if only (4) is countenanced as a sound self-predication principle for the cases in which the descriptive condition is epistemic.

5 Conclusion

In this paper I have discussed deductive systems formulated within languages containing an intensional operator ($\Box$) and a term-forming operator (ι). I have looked at two interpretations of the languages, one by Kremer and one by Horsten. The latter is intended to be a suitable interpretation for a reading of the box operator as a demonstrability or verifiability operator. Both formal interpretations are labelled ‘Carnapian’. In the Kremer interpretation variables range over individual concepts if they are within the range of the intensional operator, whereas in the Horsten interpretation variables range over individual closed terms in those circumstances. Interestingly, one cannot add one of those logics to Peano arithmetic without getting a collapse of validity or demonstrability into truth. There is however an alternative formulation of arithmetic possible which can be safely combined with any of the mentioned logics. I also looked at

[13] Actually there is a omission in the rule considered there: the exclamation mark preceding $!\phi\,(y)$ was left out. Alternatively, the exclamation mark could have been left out, but then the biconditional should have been an ordinary conditional.

certain description principles, which I dubbed 'self-predication' principles. Several among them were proved to be invalid under the Kremer interpretation. Matters were less clear if one reads them under the lights of the Horsten interpretation. At the very least one should take care in selecting a suitable elimination rule for descriptions, for otherwise one can derive the collapse of demonstrability or verifiability into truth, given some self-predication principle and the least number principle. More importantly, self-predication principles in combination with strong induction lead to a collapse of *de re* into *de dicto* verifiability. Within an arithmetical context this leads to an interesting conclusion about the limits of verifiability. The way to make progress at this point is to develop Horsten interpretations in full detail.

References

Ackerman, D. (1979). *De Re propositional attitudes toward integers. In R. Shahan and C. Swoyer (Eds.),* Essays on the Philosophy of W. V. Quine*, pp. 145–153. University of Oklahoma Press.*

Boolos, G. S., J. P. Burgess, and R. C. Jeffrey (2003). Computability and Logic *(Fourth ed.). New York: Cambridge University Press.*

Burgess, J. P. (1999). Which modal logic is the right one? Notre Dame Journal of Formal Logic 40*(1), 81–93.*

Föllesdal, D. (2004). Referential Opacity and Modal Logic*. Routledge.*

G. Hughes, M. C. (1996). A New Introduction to Modal Logic*. Routledge.*

Horsten, L. (2000). Models for the logic of possible proofs. Pacific Philosophical Quarterly 81*, 49–66.*

Horsten, L. (2005). Canonical naming systems. Minds and Machines 15*(2), 229–257.*

Horsten, L. (2008). Descriptions and the unknowable.

Kremer, M. (1997). Marti on descriptions in carnap's S2. Journal of Philosophical Logic 26*(6), 629–634.*

Marti, G. (1994). Do modal distinctions collapse in carnap's system? Journal of Philosophical Logic 23.

Shapiro, S. (Fall 2007). Classical logic. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy.