Maximally Consistent Sets of Instances of Naïve Comprehension

LUCA INCURVATI[*] & JULIEN MURZI[†]

24 May 2014

Abstract

Paul Horwich (1990) once suggested restricting the T-Schema to the maximal consistent set of its instances. But Vann McGee (1992) proved that there are multiple incompatible such sets, none of which, given minimal assumptions, is recursively axiomatizable. The analogous view for set theory—that Naïve Comprehension should be restricted according to consistency maxims—has recently been defended by Laurence Goldstein (2006; 2013). It can be traced back to W.V.O. Quine (1951), who held that Naïve Comprehension embodies the only really intuitive conception of set and should be restricted as little as possible. The view might even have been held by Ernst Zermelo (1908), who, according to Penelope Maddy (1988), subscribed to a 'one step back from disaster' rule of thumb: if a natural principle leads to contradiction, the principle should be weakened just enough to block the contradiction. We prove a generalization of McGee's Theorem, and use it to show that the situation for set theory is the same as that for

[*]University of Amsterdam and Institute for Logic, Language and Computation, L.Incurvati@uva.nl

[†]University of Kent and Munich Center for Mathematical Philosophy, Ludwig-Maximilians Universität, München, j.murzi@gmail.com

truth: there are multiple incompatible sets of instances of Naïve Comprehension, none of which, given minimal assumptions, is recursively axiomatizable. This shows that the view adumbrated by Goldstein, Quine and perhaps Zermelo is untenable.

Naïve set theory is underwritten by the *Naïve Comprehension Schema*, which consists of all instances of

(NC) $$\exists y \forall x (x \in y \leftrightarrow \Phi(x))$$

where $\Phi(x)$ is any formula in $\mathcal{L}_\in$ (the language obtained by adding '$\in$' to the language of first-order logic with identity) in which x is free and which contains no free occurrences of y. The schema is intended to express the thought that every condition determines a set: the set of all and only those objects to which the condition applies.

Naïve truth theories adhere to the unrestricted *T-Schema*:

(T) $$Tr(\ulcorner \phi \urcorner) \leftrightarrow \phi$$

where Tr is a truth-predicate and $\ulcorner \phi \urcorner$ is, for present purposes, the numerical code of ϕ relative to some fixed coding scheme. The principle trades on the putative fact that ϕ and "ϕ' is true' have the same content, and hence should be at least materially equivalent.

Naïveté has a price, however: unless classical logic is quite radically weakened, paradox looms. Let $\Phi(x)$ in (NC) be $x \notin x$, and you get Russell's Paradox. Take ϕ in (T) to be $\lambda = \neg Tr(\ulcorner \lambda \urcorner)$, and you get the Liar Paradox. This much is well-known. What is not so well-known is whether (NC) and (T) can be consistently restricted in a way that is faithful to the spirit of the naïve conceptions they characterize.

In this note, we examine the view that naïve principles should be restricted according to consistency maxims. A proposal along these lines was once made by Paul Horwich (1990) for truth, and has recently been

defended by Laurence Goldstein (2006; 2013) for both truth and set theory.

For the case of sets, the view goes back at least as far as W.V.O. Quine (1951). In his view, (NC) embodies the only really intuitive conception of set, and should, in keeping with the maxim of minimal mutilation, be restricted as little as possible so as to avoid the paradoxes. He writes:

> Only because of Russell's paradox and the like do we not adhere to the naïve and unrestricted comprehension schema [...] Having to cut back because of the paradoxes, we are well advised to mutilate no more than what may be fairly responsible for the paradoxes. (Quine 1951, pp. 50–1)

The view might even have been held at one time by Ernst Zermelo (1908, p. 263–4). According to Penelope Maddy (1988, p. 485), Zermelo saw himself as giving us as much of (NC) as possible without inconsistency. With Zermelo, Maddy suggests, it emerged the rule of thumb she labels 'one step back from disaster': if a natural principle leads to contradiction, the principle should be weakened just enough to block the contradiction.

Strictly speaking, Goldstein also requires acceptable instances of the naïve schemata not to entail tautologies, as well as contradictions.[1] However, it is easy to see that such a proposal would involve quite a radical

---

[1]Tautologies and contradictions, says Goldstein, lack content, and hence 'fail to state conditions or to give a recipe for determining whether any given object belongs in the set' (Goldstein 2013, p. 35) in the case of sets, and fail to specify coherent truth-conditions in the case of truth. Thus, $\exists y \forall x (x \in y \leftrightarrow x \notin x)$ is not an acceptable instance of (NC), and $\mathsf{Tr}(\ulcorner \lambda \urcorner) \leftrightarrow \lambda$ is not an acceptable instance of (T): both entail a contradiction, and hence are 'vacuous' and specify conditions that are 'impossible to satisfy' (Goldstein 2006, p. 886). In Goldstein's view, this is the reason for thinking that the set-theoretic and semantic paradoxes belong to the same family: '[h]erein lies the similarity between these two families of paradox. The sets and the statements we thought we were dealing with are irredeemably ill-defined, and so do no exist' (Goldstein 2006, p. 887). It is then only natural, Goldstein suggests, to restrict (NC) and (T) to instances that are neither vacuous nor impossible to satisfy. The restriction, he adds, is a principled one: it follows from the Tractarian view that logical truths and falsehoods do not express propositions (Wittgenstein 1922).

departure from classical logic.[2] We are here interested in the prospects for curtailing the Naïve Comprehension Schema whilst retaining classical logic, and focus on the idea that the problematic instances to be banned are the inconsistent ones—the ones giving rise to paradox.[3] We argue that the prospects for such a strategy are not good: restrictions of the kind Goldstein advocates fail for essentially the same reasons why, as Vann McGee (1992) first noticed, Horwich's proposed restriction of the T-Schema also fails.

1 Consistency

Goldstein writes:

> The naive comprehension axiom captures a very clear conception of sets, and it makes sense to tamper with it as little as possible, identifying and justifying the restriction that must be made to its application, rather than building up set theory from entirely new foundations. (Goldstein 2013, pp. 35–6)

So how can (NC) be made consistent, without tampering too much with the naïve conception of set?

If classical logic is to be preserved, Goldstein's idea amounts to restricting (NC) according to the following injunction:

> *Consistent*. An instance ϕ of (NC) is admissible just in case $\phi \nvdash \perp$.[4]

---

[2]In classical logic, tautologies are entailed by any sentences whatsoever, and hence also by instances of (NC). Goldstein is, of course, well aware of this, and in fact endorses a non-classical logic (see Goldstein 2012).

[3]To be sure, a different strategy for the naïve set theorist would be not to restrict (NC), and, in order to avoid triviality, *weaken the logic* instead. See e.g. Grishin 1981, Weber 2010 and Weber 2012. In the remainder of the paper, we focus on *classical* naïve but consistent set theory. We briefly return to non-classical naïve set theory in fn. 15 below.

[4]Here and throughout, '⊢' stands for first-order deducibility.

Unfortunately, this restriction will not do. The problem is that it is easy to exhibit instances of (NC) which are consistent given extremely weak assumptions—and therefore admissible according to *Consistent*—but together lead to inconsistency. We give two examples.

The first example rests on the observation that instances of (NC) can be obtained by considering formulae $\Phi(x)$ containing free variables other than x, known as *parameters*. Such instances are assumed to be prefaced by universal quantifiers binding these free variables. One such instance—in effect, also an instance of ZFC's Separation Schema—asserts, for any set, the existence of the subset of all non-self-membered members of that set. Trouble arises when we consider this instance in conjunction with one asserting the existence of a universal set.

Example 1. The following are instances of (NC):

(1) $\exists y \forall x (x \in y \leftrightarrow x = x)$

(2) $\forall z \exists y \forall x (x \in y \leftrightarrow (x \in z \ \& \ x \notin x))$

(1) has a model consisting of just one object a such that $a = \{a\}$.[5] (2) is true in any model consisting of a single hereditarily finite set.[6] The simplest such model is the one consisting of just one object a such that $a \notin a$, that is the model consisting of just the empty set—the first hereditarily finite set.[7] Hence, (1) and (2) are both admissible according to *Consistent*. But it is

---

[5]Alternatively, if one prefers a model without non-well-founded sets, one can consider a model which consists of a looped one-node graph, in which variables range over nodes, and in which $x \in y$ just in case there is a directed edge from the node assigned to y to the node assigned to x.

[6]A set is said to be *hereditarily finite* iff it is finite, its members are finite, the members of its members are finite, and so on. In other words, a set is hereditarily finite iff its transitive closure is finite.

[7]In fact, as a referee pointed out, it is clear that (2) holds in any model satisfying $\forall x (x \notin x)$. Take any set z in such a model and let y be z itself. Then (2) reduces to $\forall x (x \in z \leftrightarrow (x \in z \ \& \ x \notin x))$. But since $\forall x (x \notin x)$ holds in this model, this reduces to the logical truth $\forall x (x \in z \leftrightarrow x \in z)$.

routine to derive a contradiction from (1) and (2) together.[8]

In the light of Example 1, one might be tempted to ban instances of (NC) that are obtained using parameters. Our second example shows that this is not enough to dispense with the problem. The example makes use of the following observation, which goes back to George Boolos (1993, p. 233):

Lemma 2. *For each ϕ in $\mathcal{L}_\in$,*

(i) $\vdash \exists y \forall x (x \in y \leftrightarrow (\neg\phi \,\&\, x \notin x)) \rightarrow \phi$

(ii) $\exists y \forall x (x \notin y) \vdash \phi \rightarrow \exists y \forall x (x \in y \leftrightarrow (\neg\phi \,\&\, x \notin x))$[9]

Example 3. Take your favourite σ such that both σ and $\neg\sigma$ are consistent—for instance, let σ be shorthand for 'There are at least thirty objects'. Then, the following are instances of (NC):

(3) $\exists y \forall x (x \in y \leftrightarrow (\neg\sigma \,\&\, x \notin x))$

(4) $\exists y \forall x (x \in y \leftrightarrow (\sigma \,\&\, x \notin x))$

---

[8](1) immediately entails $\forall x (x \in a \leftrightarrow x = x)$. Instantiating (2) with a (the universal set) we have that there is a set of all non-self-membered members of a: $\exists y \forall x (x \in y \leftrightarrow (x \in a \,\&\, x \notin x))$. From this, we get $b \in b \leftrightarrow (b \in a \,\&\, b \notin b)$. But instantiating $\forall x (x \in a \leftrightarrow x = x)$ with b we also have that $b \in a \leftrightarrow b = b$. Therefore, $b \in b \leftrightarrow b \notin b$. Contradiction.

[9]For (i), suppose $\exists y \forall x (x \in y \leftrightarrow (\neg\phi \,\&\, x \notin x))$. Now suppose towards $\exists$-elimination that $\forall x (x \in a \leftrightarrow (\neg\phi \,\&\, x \notin x))$. Then we have that $a \in a \leftrightarrow (\neg\phi \,\&\, a \notin a)$. Now suppose $a \in a$. Then $\neg\phi \,\&\, a \notin a$ and so $a \notin a$. Contradiction. So $a \notin a$. But then suppose $\neg\phi$. Then $a \notin a \,\&\, \neg\phi$ and so $a \in a$. Contradiction again. By classical reductio it follows that ϕ. But ϕ has been obtained on the supposition that $\forall x (x \in a \leftrightarrow (\neg\phi \,\&\, x \notin x))$. Hence, by $\exists$-elimination we have ϕ on the overall supposition that $\exists y \forall x (x \in y \leftrightarrow (\neg\phi \,\&\, x \notin x))$. The result follows by conditional proof.

For (ii), first notice that $b \notin a$ follows from our premiss $\exists y \forall x (x \notin y)$ and suppose ϕ. Now suppose $b \in a$. This contradicts $b \notin a$, and so we can infer $\neg\phi \,\&\, b \notin b$. Next, suppose $\neg\phi \,\&\, b \notin b$ and further assume $b \notin a$ for $\neg$-introduction. Since both ϕ and $\neg\phi$ are among our assumptions, $b \in a$ follows by classical reductio. We have derived $\neg\phi \,\&\, b \notin b$ from $b \in a$ and $b \in a$ from $\neg\phi \,\&\, b \notin b$. By $\leftrightarrow$-introduction, it follows that $b \in a \leftrightarrow (\neg\phi \,\&\, b \notin b)$. Applying $\forall$-introduction and $\exists$-introduction, we obtain $\exists y \forall x (x \in y \leftrightarrow (\neg\phi \,\&\, x \notin x))$. Thus, we have $\exists y \forall x (x \in y \leftrightarrow (\neg\phi \,\&\, x \notin x))$ on the overall supposition that ϕ. The result follows by conditional proof.

By Lemma 2(ii), any proof of a contradiction from (3) could be transformed, in the presence of the Empty Set Axiom, into a proof of a contradiction from σ, and similarly for (4). Hence, in the presence of the Empty Set Axiom, (3) and (4) are both admissible instances of (NC) according to *Consistent*. But by Lemma 2(i), (3) entails σ, and (4) entails $\neg\sigma$. Contradiction.

In short: instances of (NC) that are admissible according to *Consistent* may be *jointly* inconsistent, and hence unacceptable.[10] This suggests that naïve but consistent set theorists must focus on acceptable *sets* of instances of (NC)—a suggestion to which we now turn.

2   Maximal consistency

The naïve set theorist might take the results of the previous section to simply show that, whenever the addition of a consistent instance of (NC) yields an inconsistency, the instance is not acceptable after all. That is, she might take the naïve principles to be so intuitive that we should keep as many instances of them as we consistently can. More formally, the acceptable instances of (NC) must form a *maximally consistent* set:

---

[10]The situation here is analogous to that arising in the context of the debate on the acceptability of *abstraction principles*. These are principles of the form

$$\S(\alpha) = \S(\beta) \leftrightarrow \alpha \approx \beta$$

where § is an operator taking items of type α and β to be objects, and $\approx$ is an equivalence relation on items of that type. In *Grundlagen* (1884), Frege considered and discarded the possibility of using a principle of this sort to contextually define the concept of number. And he later took extensions to be governed by an abstraction principle stating that two concepts have the same extension just in case they are coextensive. As is well-known, the principle concerning extensions is inconsistent, so any attempt to defend the legitimacy of using abstraction principles to define mathematical concepts (such as Wright 1983) must address the question of which abstraction principles are acceptable to this end. But—and here lies the analogy with the case of (NC)—it will not do simply to say that the acceptable abstraction principles are the consistent ones. For there are abstraction principles that are individually consistent but jointly inconsistent (see e.g. Boolos 1990). For more on the acceptability of abstraction principles, see Fine 2002.

Maximally Consistent. A set of instances Γ of (NC) is admissible just in case $\Gamma \nvdash \bot$ and if Λ is a set of instances of (NC) and properly extends Γ, then $\Lambda \vdash \bot$.[11]

In fact, this proposal seems to better capture the spirit of Quine's remark quoted above, and seems more in keeping with the one step back from disaster rule of thumb, which, if Maddy is right, the early Zermelo subscribed to. The approach was once advocated by Horwich for truth:

[I]t suffices for us to concede that certain instances of [the T-Schema (T)] are not to be included as axioms of the minimal theory, and to note that the principles governing our selection of excluded instances are, in order of priority: (a) that the minimal theory not engender 'liar-type' contradictions; (b) that the set of excluded instances be as small as possible; and [...] (c) that there be a constructive specification of the excluded instances that is as simple as possible. (Horwich 1990, p. 42)

The view faces a twofold problem, however, as McGee (1992) first pointed out. First, in the presence of a theory strong enough to interpret Robinson's Arithmetic (henceforth Q), maximally consistent sets of instances of (T) are not recursively axiomatizable, and hence it is unclear whether they can be said to be *theories* of truth. Second, there exist several maximally consistent, and yet incompatible, sets of instances of (T). We are effectively presented with an embarrassment of riches, and it would seem arbitrary to choose one over another. We now show that the suggestion that (NC) must be restricted according to *Maximally Consistent* faces exactly the same difficulties.

---

[11]As usual, we say that a set Λ *extends* a set Γ iff Γ is a subset of Λ, and that it *properly extends* Γ iff Γ is a proper subset of Λ.

3 Generalizing McGee's result

We want to show that McGee's result admits of a natural generalization. To this end, we first introduce some notation and terminology. If T is a theory, we write '$\Gamma \vdash_\mathsf{T} \phi$' for '$\Gamma \cup \mathsf{T} \vdash \phi$'. And we make use of the following definitions:

Definition 4. Let Γ be a set of sentences and Σ a first-order schema.

   (i) Γ is T-*consistent* iff $\Gamma \nvdash_\mathsf{T} \bot$.

  (ii) Γ is Σ-*maximally* T-*consistent* iff Γ is T-consistent and any set which properly extends Γ and contains Σ-instances not in Γ is T-inconsistent.

 (iii) Γ is *maximally* T-*consistent* iff it is T-consistent and any set which properly extends Γ is T-inconsistent.

Clearly, this is an increasing order: if Γ is maximally T-consistent, then it is Σ-maximally T-consistent; and if Γ is Σ-maximally T-consistent, then it is T-consistent.

   Now let S be a consistent first-order theory entailing the axioms of some theory S', let Σ be a first-order schema, and let $\mathcal{L}$ be a countable first-order language. Suppose we relatively interpret in $\mathcal{L}$ the language obtained by adding to the language of S any non-logical expression in Σ. We can then prove the following:

Theorem 5. *Suppose that for each $\phi \in \mathcal{L}$ there is an instance Σ_ϕ of Σ such that*

$$\vdash_{\mathsf{S}'} \phi \leftrightarrow \Sigma_\phi$$

Then:

   (1) For any S-consistent set Δ of sentences in $\mathcal{L}$, there is a Σ-maximally S-consistent set Γ of Σ-instances such that $\Gamma \vdash_\mathsf{S} \delta$ for every $\delta \in \Delta$.

(2) If Ξ is a Σ-maximally S-consistent set of Σ-instances, then, for every $\psi \in \mathcal{L}$,
$\Xi \vdash_{S'} \psi$ *or* $\Xi \vdash_{S'} \neg\psi$.

Proof. For (1), let $\Theta = \{\Sigma_\delta : \delta \in \Delta\}$. Note that $\Theta \vdash_S \delta$ for every $\delta \in \Delta$, since

$$\vdash_S \delta \leftrightarrow \Sigma_\delta$$

for every $\delta \in \Delta$. Moreover, since Δ is S-consistent, so is Θ. Hence, by Zorn's Lemma, Θ can be extended to a Σ-maximally S-consistent set Γ of Σ-instances (see Lemma 8 in Appendix A for details).

For (2), let Ξ be a Σ-maximally S-consistent set of Σ-instances, and take a $\psi \in \mathcal{L}$. Since Ξ is S-consistent, then either $\Xi \cup \{\psi\}$ or $\Xi \cup \{\neg\psi\}$ is S-consistent.[12] So either $\Xi \cup \{\Sigma_\psi\}$ or $\Xi \cup \{\Sigma_{\neg\psi}\}$ is S-consistent. But since Ξ is Σ-maximally S-consistent, either $\Sigma_\psi \in \Xi$ or $\Sigma_{\neg\psi} \in \Xi$. Thus, $\Xi \vdash_{S'} \psi$ or $\Xi \vdash_{S'} \neg\psi$. $\square$

Note that although the first part of the theorem only guarantees the existence of a certain Σ-maximally S-consistent set Γ of Σ-instances, it follows from the second part that a corresponding maximally S-consistent set can be obtained simply by closing Γ under $\vdash_S$.[13] Next, note that we could take S to be a theory which interprets Q. It would then follow from the second part of the theorem that the theory obtained by adding any Σ-maximally

---

[12]For suppose Ξ is S-consistent but neither $\Xi \cup \{\psi\}$ nor $\Xi \cup \{\neg\psi\}$ is S-consistent. From the fact that $\Xi \cup \{\psi\}$ is S-inconsistent it follows that $\Xi \cup \{\psi\} \vdash_S \neg\psi$ and so, by the deduction theorem, that $\Xi \vdash_S \psi \to \neg\psi$. But since $(\psi \to \neg\psi) \to \neg\psi$ is a first-order validity and hence $\Xi \vdash_S (\psi \to \neg\psi) \to \neg\psi$, we have that $\Xi \vdash_S \neg\psi$. A similar reasoning can be used to show that $\Xi \vdash_S \psi$ follows from the fact that $\Xi \cup \{\neg\psi\}$ is S-consistent. But then $\Xi \vdash_S \neg\psi$ and $\Xi \vdash_S \psi$, contradicting the initial supposition that Ξ is S-consistent.

[13]For let Γ be a Σ-maximally S-consistent set of Σ-instances and let $\mathrm{Ded}_S(\Gamma)$ be the set of all ϕs such that $\Gamma \vdash_S \phi$. Clearly, $\mathrm{Ded}_S(\Gamma)$ is S-consistent. Now suppose there is a set Λ which properly extends $\mathrm{Ded}_S(\Gamma)$. Then, there is some $\phi \in \mathcal{L}$ such that $\phi \notin \mathrm{Ded}_S(\Gamma)$ but $\phi \in \Lambda$. Since Γ is a Σ-maximally S-consistent set of Σ-instances, it follows from the second part of Theorem 5 that for every $\psi \in \mathcal{L}$ either $\psi \in \mathrm{Ded}_S(\Gamma)$ or $\psi \notin \mathrm{Ded}_S(\Gamma)$. So $\neg\phi \in \mathrm{Ded}_S(\Gamma)$. But since Λ is an extension of $\mathrm{Ded}_S(\Gamma)$, this means that $\neg\phi \in \Lambda$ too. Thus, Λ is S-inconsistent.

S-consistent set of Σ-instances to S is negation complete, and hence non-recursively axiomatizable.[14] Finally, although we have used Zorn's Lemma in the proof of (1), this could be dispensed with by constructing Γ step by step (see Appendix A for details).

Given the above generalization of McGee's proof, we can obtain the desired result in the case that interests us by using Boolos's observation mentioned above. (McGee's result can be recovered as a special case of Theorem 5, as we show in Appendix B below.) Let R be some consistent theory entailing the axioms of the theory whose sole non-logical axiom is

(NC$_\emptyset$) $\qquad\qquad\qquad \exists y \forall x (x \in y \leftrightarrow x \neq x)$

And suppose that we have relatively interpreted the language of R in a countable first-order language $\mathcal{L}'$. We have:

Lemma 6. *For each $\phi \in \mathcal{L}'$,*

$$\vdash_{\mathsf{NC}_\emptyset} \phi \leftrightarrow \exists y \forall x (x \in y \leftrightarrow (\neg\phi \,\&\, x \notin x))$$

Proof. Immediate from Lemma 2. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

But together with Theorem 5, Lemma 6 immediately yields:

Corollary 7.

(1) For any R-consistent set Δ of sentences in $\mathcal{L}'$, there is a (NC)-maximally R-consistent set Γ of (NC)-instances such that $\Gamma \vdash_{\mathsf{R}} \delta$ for every $\delta \in \Delta$.

(2) If Ξ is a (NC)-maximally R-consistent set of (NC)-instances, then, for every $\psi \in \mathcal{L}'$, $\Xi \vdash_{\mathsf{NC}_\emptyset} \psi$ or $\Xi \vdash_{\mathsf{NC}_\emptyset} \neg\psi$.

---

[14]That such a theory cannot be recursively axiomatized follows from J. Barkley Rosser's (1936) strengthening of Gödel's (1931) First Incompleteness Theorem, which establishes that no consistent recursively axiomatizable theory interpreting Q is negation complete.

4 Maximal consistency is not enough

Corollary 7 has two key philosophical consequences. First, take any theory which interprets Q and contains a set of instances of (NC) which includes $NC_\varnothing$ and is (NC)-maximally consistent with the rest of the theory. The second part of the Corollary tells us that any such theory is negation complete, and hence cannot be recursively axiomatized. Clearly, however, any set theory worth its name should interpret Q. Moreover, $NC_\varnothing$, which simply asserts the existence of the empty set, should be an acceptable instance of (NC) if anything is. Thus, there is no reasonably strong recursively axiomatizable theory that satisfies the injunction dictated by *Maximally Consistent*: to identify our theory of sets with all instances of (NC) that are admissible according to *Maximally Consistent* would seem to fall short of providing a set *theory*.

Second, even setting this problem aside, the proposed emendation of naïve set theory is fundamentally flawed. The first part of Corollary 7 gives us a recipe for constructing several admissible, and yet incompatible, sets of instances of (NC), each of which has an equal claim to embody the naïve but consistent conception of set on offer.

To wit, take a σ such that σ and $\neg\sigma$ are (individually) consistent with the existence of the empty set. Then, start with the sets $\{\exists y \forall x (x \in y \leftrightarrow (\neg\sigma \ \& \ x \notin x)), NC_\varnothing\}$ and $\{\exists y \forall x (x \in y \leftrightarrow (\sigma \ \& \ x \notin x)), NC_\varnothing\}$, and use Corollary 7 to extend each of these sets to a (NC)-maximally consistent set of (NC)-instances. The resulting two (NC)-maximally consistent sets are incompatible given Lemma 6: taking ϕ as σ in Lemma 6, we get the result that the first set implies σ, while taking ϕ as $\neg\sigma$ we have the result that the second set implies $\neg\sigma$. In order to restore consistency, the naïve but consistent set theorist must settle on one such (NC)-maximally consistent set, as embodying her conception. Yet, we are given no guidance at all as to how this choice is to be made: consistency was meant to be our sole guide for restricting (NC).

To put matters in different terms, the naïve but consistent set theorist adopts the one step back from disaster rule of thumb, which urges us to restrict (NC) just enough to avoid inconsistency. However, there are several, incompatible directions in which such a step back could be taken. Absent some other rule of thumb, we are at a loss as to what direction should be favoured. We are inclined to interpret the provision of several mutually incompatible tamperings of the naïve conception of set as the provision of no conception at all. If our logic is to remain classical, or classical enough, yet more tinkering with (NC) is required.[15]

5   Concluding remarks

The semantic and set-theoretic paradoxes are standardly taken to show that the naïve conceptions of set and truth are untenable, at least to the extent that we want to hold on to classical logic. Our results show that more is in fact true: such conceptions remain untenable even if, following the likes of Horwich and Quine, we give up precisely the portion of naïveté that generates paradox. If our logic is to remain classical, a more radical departure from the naïve schemata is required. This is the strategy pursued, for instance, by groundedness approaches (Leitgeb 2005) and revision theories in the case of truth (Gupta and Belnap 1993), and by the iterative conception (Boolos 1971) and the graph conception (Incurvati 2014) in the case of sets. Our results do not adjudicate between these approaches, and leave it open which of them, if any, should be favoured. Still, they do

---

[15]Another option would be to both restrict (NC) according to consistency maxims *and* substantially weaken classical logic. We do not know whether our results would then carry over: they hold for classical logical consequence, but they need not continue to hold for weaker consequence relations. As mentioned above, Goldstein himself endorses a non-classical logic and hence our results need not undermine his proposal. Here we simply notice that the suggestion of weakening the logic would be open to the potential problem that a set theory that is not strong enough to validate the results of this paper might be too weak. In any event, pending a precise description of the logic Goldstein favours, the issue remains highly speculative.

suggest that *some* approach of this kind is called for. Our final diagnosis is not different from McGee's: we take our results to be evidence that restrictions of the naïve set-theoretic and semantic principles need to be guided by *non-naïve* conceptions of sets and truth.[16]

Appendix A

We first prove the lemma (a version of Lindebaum's Lemma) used in the proof of part (1) of Theorem 5:

Lemma 8. *Each S-consistent set of Σ-instances can be extended to a Σ-maximally S-consistent set of Σ-instances.*

Proof. Suppose Ξ is a S-consistent set of Σ-instances. Let P be the set of all S-consistent sets of Σ-instances extending Ξ, partially ordered by inclusion $\subseteq$. If $\mathcal{C}$ is a chain of elements in P, then $\bigcup \mathcal{C}$ is S-consistent. So $\bigcup \mathcal{C}$ is in P, since $\bigcup \mathcal{C}$ contains every S-consistent set of Σ-instances extending Ξ. By Zorn's Lemma, P has a maximal element Γ. Now suppose that Γ is not a maximal element of P. Then, there is some $\psi \in \mathcal{L}$ such that $\Gamma \nvdash_S \Sigma_\psi$ and $\Gamma \nvdash_S \Sigma_{\neg\psi}$. If $\Gamma \nvdash_S \Sigma_\psi$ then $\Sigma_\psi \notin \Gamma$. And if $\Gamma \nvdash_S \Sigma_{\neg\psi}$, then $\Gamma \cup \{\Sigma_\psi\}$ is S-consistent and therefore in P. So $\Gamma \cup \{\Sigma_\psi\}$ is an S-consistent set of Σ-instances that properly extends Γ, contradicting the fact that Γ is a maximal element of P. Therefore, Γ is Σ-maximally S-consistent. $\square$

We then prove part (1) of Theorem 5 without using Zorn's Lemma:

Proof. Enumerate the sentences of $\mathcal{L}$ as $\psi_0, \psi_1, \psi_2, \ldots$ and define an increasing sequence $\Gamma_0, \Gamma_1, \Gamma_2, \ldots$ of finite sets of Σ-instances as follows:

$$\Gamma_0 = \{\Sigma_\delta : \delta \in \Delta\}$$

$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{\Sigma_{\psi_n}\} & \text{if } \Gamma_n \cup \{\psi_n\} \text{ is S-consistent} \\ \\ \Gamma_n \cup \{\Sigma_{\neg\psi_n}\} & \text{otherwise} \end{cases}$$

Now let $\Gamma = \bigcup_{n \geq 0} \Gamma_n$. Clearly, $\Gamma \vdash_S \delta$ for every $\delta \in \Delta$. Moreover, Γ_0 is S-consistent if Δ is, and the sequence $\Gamma_0, \Gamma_1, \Gamma_2, \ldots$ is defined so that if Γ_n is S-consistent, then Γ_{n+1} is S-consistent too. By induction on n, it follows that if Δ is S-consistent, then so is Γ_n for each n, and hence Γ itself.

To check that Γ is a Σ-maximally S-consistent set of Σ-instances, take $\psi \in \mathcal{L}$ and suppose $\Sigma_\psi \notin \Gamma$. ψ must appear somewhere in the list $\psi_0, \psi_1, \psi_2, \ldots$, that is $\psi = \psi_m$ for some m. But if $\Sigma_{\psi_m} \notin \Gamma$, this must be because $\Gamma_m \cup \{\psi_m\}$ is S-inconsistent. And this means that $\Gamma_m \cup \{\Sigma_{\psi_m}\}$ is S-inconsistent, since

$$\vdash_S \psi_m \leftrightarrow \Sigma_{\psi_m}$$

Since $\Gamma_m \subseteq \Gamma$, $\Gamma \cup \{\Sigma_\psi\}$ is S-inconsistent too. $\qquad\square$

Appendix B

For completeness, we show how McGee's result can be recovered as a special case of Theorem 5. In this case, we work with a countable first-order language $\mathcal{L}^\star$ in which we have relatively interpreted the language obtained by adding Tr to the language of arithmetic. We make use of the following lemma:

Lemma 9 (McGee)**.** *For each ϕ in $\mathcal{L}^\star$, there is a sentence B_ϕ such that*

$$\vdash_Q \phi \leftrightarrow (\mathsf{Tr}(\ulcorner B_\phi \urcorner) \leftrightarrow B_\phi)$$

Proof. Apply Tarski's Fixed-Point Theorem to the $\mathcal{L}^\star$-predicate $(\phi \leftrightarrow \mathsf{Tr}(x))$, and then use the associativity of biconditionals. □

Now let A be a consistent arithmetical theory which implies the axioms of Q. Then, in conjunction with Theorem 5, Lemma 9 immediately delivers McGee's result:

Corollary 10 (McGee)**.** *For any A-consistent set Δ of sentences in $\mathcal{L}^\star$, there is a (T)-maximally A-consistent set Γ of (T)-instances such that $\Gamma \vdash_A \delta$ for every $\delta \in \Delta$. Moreover, for all $\psi \in \mathcal{L}^\star$, $\Gamma \vdash_Q \psi$ or $\Gamma \vdash_Q \neg\psi$.*

References

Boolos, George 1971: 'The iterative conception of set', in his 1998, pp. 13–29. Originally published in *Journal of Philosophy*, 68, 1971, pp. 215–31.

Boolos, George (ed.) 1990a: *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge: Cambridge University Press.

Boolos, George 1990b: 'The standard of equality of numbers', in his 1998, pp. 202–219. Originally published in Boolos 1990a, pp. 261–77.

Boolos, George 1993: 'Whence the contradiction?', in his 1998, pp. 220–36. Originally published in *Proceedings of the Aristotelian Society (supplementary volume)*, 67, 1993, pp. 213–33.

Boolos, George 1998: *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Fine, Kit 2002: *The Limits of Abstraction*. Oxford: Oxford University Press.

Frege, Gottlob 1884: *Die Grundlagen der Arithmetik*. Breslau: W. Koebner.

Gödel, Kurt 1931: 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I'. *Monatshefte für Mathematik und Physik*, 38, pp. 173–98.

Goldstein, Laurence 2006: 'Fibonacci, Yablo, and the cassationist approach to paradox'. *Mind*, 115, pp. 867–89.

Goldstein, Laurence 2012: 'Reasoning without contradiction'. *The Reasoner*, 6, pp. 183–4.

Goldstein, Laurence 2013: 'Paradoxical partners: semantical brides and set-theoretical grooms'. *Analysis*, 73, pp. 33–7.

Grishin, Vyacheslav Nikolaevich 1981: 'Predicate and set-theoretic calculi based on logic without contractions'. *Mathematics of the USSR-Izvestiya*, 45, pp. 45–68.

Gupta, Anil and Nuel Belnap 1993: *The Revision Theory of Truth*. Cambridge, MA: MIT Press.

Horwich, Paul 1990: *Truth*. Oxford: Oxford University Press, second edition 1998.

Incurvati, Luca 2014: 'The graph conception of set'. *Journal of Philosophical Logic*, 43, pp. 181–208.

Leitgeb, Hannes 2005: 'What truth depends on'. *Journal of Philosophical Logic*, 34, pp. 155–92.

Maddy, Penelope 1988: 'Believing the axioms. I'. *Journal of Symbolic Logic*, 53, pp. 481–511.

McGee, Vann 1992: 'Maximal consistent sets of instances of Tarski's schema (T)'. *Journal of Philosophical Logic*, 21, pp. 235–41.

Quine, W. V. O. 1951: *Set Theory and Its Logic*. Cambridge, MA: Belknap Press.

Rosser, J. Barkeley 1936: 'Extensions of some theorems of Gödel and Church'. *Journal of Symbolic Logic*, 1, pp. 87–91.

Weber, Zach 2010: 'Transfinite numbers in paraconsistent set theory'. *Review of Symbolic Logic*, 3, pp. 71–92.

Weber, Zach 2012: 'Transfinite cardinals in paraconsistent set theory'. *Review of Symbolic Logic*, 5, pp. 269–93.

Wittgenstein, Ludwig 1922: *Tractatus Logico-Philosophicus*. London: Kegan Paul.

Wright, Crispin 1983: *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

Zermelo, Ernst 1908: 'Untersuchungen über die Grundlagen der Mengenlehre I'. *Mathematische Annalen*, 65, pp. 261–81.