Review of: “Consequences of an exotic definition for $P = NP$” by N.C.A. da Costa and F.A. Doria

by Ralf Schindler

Institut für Mathematische Logik und Grundlagenforschung
Westfälische Wilhelms-Universität Münster, Einsteinstr. 62, 48149 Münster, Germany

`rds@math.uni-muenster.de`

To any first order theory T which contains a reasonable fragment of Peano Arithmetic we may associate computable functions $f\colon \mathbb{N} \to \mathbb{N}$ such that f is total in reality but T cannot prove that f be total. For instance, we may let $f_T(n)$ be $\sigma + 1$ where σ is the maximum of all values $\{e\}(k)$ such that $k \leq n$ and T proves that $\{e\}$ is a total function from $\mathbb{N}$ to $\mathbb{N}$ via a proof of length $\leq n$. The paper under review provides a failing attempt to exploit this phenomenon for showing that $P = NP$ be consistent with the system ZFC of set theory.

An ealier such attempt was “On the consistency of $P = NP$ with fragments of ZFC whose own consistency strength can be measured by an ordinal assignment” by the same authors (available at `xxx.lanl.gov/abs/math.LO/0006079`). However, this attempt failed as well; a criticism may be found in “A note on an alleged proof of the relative consistency of $P = NP$ with PA” by the present reviewer (available at `xxx.lanl.gov/abs/math.LO/0007025`).

What the paper currently under review shares with the above-cited older one is that it only makes use of pretty elementary methods and that nevertheless crucial steps in the argument are not spelled out at all; in both cases, one such step just appears to be a gap for which no filling is in sight.

Let T be the theory ZFC, and let $f = \{e\}$ be a computable function, where e is a concrete integer. The authors define a Σ^0_2 sentence of the language of ZFC which they denote by $[P = NP]^f$ and which says that either $\{e\}$ is not total or else there is some integer m and a Turing machine with running time $n \mapsto O(n^{\{e\}(m)})$ (sic!) which decides a given NP-complete problem (the satisfyability problem for formulas of sentential logic, say). If id denotes the identity then $[P = NP]^{\mathrm{id}}$, abbreviated by $[P = NP]$, is a fair formal version of the statement that P equals NP.

If $f = f_T = f_{ZFC}$ is the function defined in the first paragraph above then it is less clear if $[P = NP]^f$ should count as a fair version of the statement that $P = NP$, but anyway if a strengthening of ZFC proves that f is total then it also proves that $[P = NP] \longleftrightarrow [P = NP]^f$.

The alleged key results of the paper are supposed to be proven on pages 662 and 663 culminating in Corollary 4.6 which asserts that $[P = NP]$ is relatively consistent with ZFC. By what we have discussed so far it is clear that in order to verify Corollary 4.6 it would be enough to prove that for some $f = \{e\}$, $ZFC + [P = NP]^f +$ "f is total" is consistent, assuming the consistency of ZFC; to try and prove this is an ambitious challenge. Nevertheless, this is exactly what the authors pretend on having achieved by their "informal proof" of Proposition 4.4; it is here (on p. 662 lines 5 to 3 from bottom) where the crucial gap of the paper appears.

It is a short "proof" indeed for such a strong statement. If ZFC is consistent then so is $ZFC + [P = NP]^f$, just because ZFC can not prove that f be total (cf. the proof of Proposition 4.1). An appeal to the possibility of further adding a reflection principle to $ZFC + [P = NP]^f$ without introducing an inconsistency is then used for getting the desired result. This appeal remains a complete mystery. It is true that there are reflection principles which can be added to ZFC alone without introducing an inconsistency, but it is entirely dubious why this should also be possible for $ZFC + [P = NP]^f$. In fact, if ZFC were to prove $\neg[P = NP]$ (a widely held belief) then such a possibility would be ruled out.

In the author's model of $ZFC + [P = NP]^f$, a statement which they denote by $[ZFC\ is\ \Sigma_1 sound]$ *fails* by their Lemma 4.3; $[ZFC\ is\ \Sigma_1 sound]$ is supposed to follow from a reflection principle; therefore, that reflection principle also has to *fail* in their model of $ZFC + [P = NP]^f$. The conclusion should therefore really have been: said reflection principle is *not* relatively consistent with $ZFC + [P = NP]^f$.