Page 5, last paragraph, and page 6, first paragraph: I am at a loss to understand the main point of these two paragraphs. For instance, what do you mean by automated theorem provers achieving a part of the ambitions suggested by Hilbert in **? How does a prover "clarify the nature of the infinite" or achieve a similar goal? The following paragraph abruptly switches from automated provers to humans, stating that "a human's faith in his own consistency is an essential prerequisite to gain the needed psychological motivation for stimulating cogitation" (an apparently accidental "?" appears at the end). Fine, but what does this have to do with automated provers, or with Gödel-style un*provability* of consistency?

Page 6, the long definition of $\mathrm{IS}_D(A)$: my comments concerning page 4 lines (-10)–(-8) also apply here. Moreover, I am slightly worried that none of the axiom groups contains axioms specifying the meaning of the symbols of L^* not mentioned in group zero (such as e.g. Logarithm and Root).

Page 7, item (ii.): either both x and y *are* sentences or both *represent* sentences.

Page 8, Definition 5.1: is "axioms that have Π_1^* encodings" a complicated way of saying "Π_1^* axioms"?

Page 8 line -8: what is "Π_1^* styled" information?

Page 8, Definition 5.3: unlike "kernelized formula", which is a useful concept with a rather odd name, "kernel-list" seems to be a *useless* concept. This is strongly suggested by the fact that there are no effectivity (or any other) conditions on the enumeration of the kernels (is "kernel" the same thing as "kernelized formula"?), and then confirmed by the actual use of kernel lists, which typically takes the form "for any index i, ..." This is just an unnecessarily oblique way of saying "for any kernelized formula, ...". I suggest getting rid of the whole concept of kernel list and modifying various statements accordingly.

Page 9, Example 5.4: unless you make stronger assumptions on the formula Probe(g,x), I see no reason why equivalence (6) should be provable in PA, rather than just true in the standard model.

Page 9, Theorem 5.7: I assume you want $IS_D^{\#}(\beta_{A,i})$ to be consistent. Otherwise, taking $0 \neq 0$ for $\beta_{A,i}$ would work.

Page 10 line 10: "consistent". With what? Do you want β to be simply consistent as a set of first-order sentences or consistent with some minimal theory of arithmetic?

Page 11 line 4: what does "formally true" mean?

Page 11, Remarks 6.1 and 6.2: of course, one may speculate and conjecture, but I have to say that I find the suggestions made in these remarks vauge and not too convincing. Firstly, why should we expect self-justification of the sort studied in this paper, especially with respect to very specific (and weak) proof systems, to have any importance from an engineering perspective? (A similar comment applies to the remarks on page 18 following Theorem 10.1). Secondly, why should using versions of (11) for many different truth definitions be beneficial in comparison with using (11) for a single well-chosen definition?

As context, let me give one example in which I would argue that a slightly different self-justification phenomenon *is* relevant, if not to engineering directly, then at least to theoretical work on the border of computational practice. For reasons related to Pudlák's upper bounds on the size of proofs of finitistic consistency statements, a typical *propositional* proof system has short proofs of tautologies expressing the consistency of the system. In a 2003 paper, Atserias and Bonet proved that this phenomenon extends down almost but not quite to the resolution proof system, which is actually extensively used in the practice of SAT-solving. As an application, they proved results suggesting that resolution is unlikely to have a potentially very useful property known as weak automatizability. Thus, results that point to potential limitations of a large class of SAT solvers were obtained using theoretical knowledge about self-justification-like properties of some propositional proof systems.

Page 11 line -3: "we have described our results..." Where?

Page 12 line -8: "a Yes-or-No answer to Question * * *". The question begins with "How".

Page 12 line -6: "sequence *of* integers".

page 12 line -2: "greater understanding for the statements * and **". While one might speculate about a connection, however strained, between the weak self-justifying systems and Gödel's statement * (though I think the conjectured connections of * to Gentzen- or Dialectica-style approaches are more convincing), the connection between such systems and ** seems more dubious. The sentences quoted as **, written well