Experimental logics, mechanism and knowable consistency
(SELLC 2010 in Guangzhou, China)*

Martin Kaså, University of Gothenburg, Sweden

Overview  In a paper published in 1975, Robert Jeroslow introduced the concept of an experimental logic as a generalization of ordinary formal systems such that theoremhood is a $\Pi^0_2$ property (or in practise $\Delta^0_2$) rather than $\Sigma^0_1$. These systems can be viewed as (rather crude) representations of axiomatic theories evolving stepwise over time. Similar ideas can be found in papers by Putnam (1965) and McCarthy and Shapiro (1987).

The topic of the present paper is a discussion of a suggestion by Allen Hazen, that these experimental logics might provide an illuminating way of representing “the human mathematical mind”. This is done in the context of the well-known Lucas-Penrose thesis. Though we agree that Jeroslow’s model has some merit in this context, and that the Lucas-Penrose arguments certainly are less than persuasive, some semi-technical doubts are raised concerning the alleged impact of experimental logics on the question of knowable self-consistency.

A teaser  Gödel’s incompleteness theorems have inspired some thinkers – among them Gödel himself – to draw philosophical conclusions regarding the (im)possibility of a true mechanical model of mind. The thesis is, vaguely put, that not even the arithmetical faculties of the human mind (in some sense) can be (in some sense) a machine (in Turing’s sense). The two most well-known and widely discussed such theorists are J.R. Lucas and R. Penrose who have proposed what are often lumped together as the Lucas-Penrose thesis and the Lucas-Penrose arguments. They are also known for sticking to their position in the face of an impressive host of harshly critical logicians and philosophers.

*Extended abstract submitted for the Student Session at the Sino-European Winter School in Logic, Language and Information. The area is Logic. This work is part of the author’s PhD project. REVISED AFTER REFEREE COMMENTS
The starting point of it all is Lucas’ claim that “Gödel’s theorem states that in any consistent system which is strong enough to produce simple arithmetic there are formulas which cannot be proved in the system, but which we can see to be true”, to which the obvious standard objection is that all we know for certain is that if the system \( S \) is consistent then (e.g.) \( S \nvdash \text{Con}(S) \). But, of course, (i) there is no stopping that claim from being a theorem of \( S \), and (ii) there is, in general, no reason to assume that “we” know that \( S \) in fact \textit{is} consistent. The extensive debate from there on should be well-known in its essentials to most readers, and we will not reiterate the details here.

There is absolutely no point in citing authorities and agreeing to disagree with Lucas and Penrose, but, as it happens, there is a perfectly valid reason to revisit this problem once more.

Allen Hazen and Stewart Shapiro have, independently and in different contexts, suggested a novel perspective on the debate between mechanism and anti-mechanism.

In Hazen’s case, the suggestion comes in the humble guise of two thought-provoking postings to the Foundations of Mathematics mailing-list, where he adopts the critical majority view, but, at least partly, for new reasons.

Initially, he observes that a difficulty in discussing “the Lucas-Penrose argument” is that there is no such thing – the implied uniqueness condition is not satisfied. Therefore, he sets out what he calls his “favorite version” of Lucas’ position (the wording has been changed in minor ways):

1. If the mind is mechanical, human mathematics is the product of a machine.
2. The product of a machine is a computably enumerable set, i.e. the set of theorems of some formal theory.
3. No arithmetically adequate, consistent, formal theory has a theorem asserting the consistency of that selfsame theory (Gödel).
4. But human mathematicians can know that their mathematics is (or will be) consistent.
5. Hence: The mind is not mechanical.

The idea would be to undermine step (2) by an alternate conception of “the product of a machine” in such a way that (4) is applicable to such a machine to the same extent as to the human mathematical mind. Then
Gödel’s incompleteness theorems would not be (directly) relevant, and (5) would (hopefully) not follow.

So, what are we modelling, anyway? What kind of enterprise is it that the “mechanical mathematician” shall represent? It should be recognizable as (idealized) human mathematical activity, and as such it must be fallible. Postulating that human mathematics will ever enter into a final golden era when all inconsistencies are forever exorcized seems too much of an idealization. However, it might be more justifiable to state that individual errors are spotted and corrected as time goes by. This is in line with the discussion in the expository (Shapiro 1998), where, while discussing idealizations, Shapiro contrasts the “euclidean” model (gap free deduction from self-evident axioms) with a more realistic and dynamic “semi-euclidean” model: the ability (in principle) to recognize and withdraw contradictions and other arithmetic falsehoods. So, a picture of momentarily fallible, but dynamic and (in a sense) “eventually infallible” mathematical activity might look something like the following:

Consider an ideal mathematician engaged in developing an axiomatic theory. She might start with some base theory (say first order PA) which she does not question, and then she wants to extend it by adding new concepts and new pieces of information. Adding new concepts, in a formal setting, amounts to adding new symbols to the language and adding axioms that characterize the intuitive mathematical ideas. Then, there is the business of deduction. Theorems pile up, and unwanted consequences may surface: contradictions, as well as formal theorems that show that the axiomatization fails to capture the intended informal concept. Luckily, our careful mathematician has kept track of which axioms were used in which proofs, and therefore she is in a position to backtrack, delete whatever axioms she holds responsible (as well as dependent theorems) and start anew. And thus the next step in the theory development is taken.

In a more fortunate case, she might be happy with the results so far, but still want to go on developing her theory by adding new axioms (concerning new or old non-logical symbols). Ideally, this process continues without limitations of space, time, etc. Not caring too much about what happens at each step, nor about why and how new axioms are chosen, we get a “global” picture of a discrete theory development over time. And the theory as a whole, the dynamic theory, is this entire sequence. Now, there seems to be no prima facie reason to believe that the “set of mathematical theorems” (or, perhaps better: “the humanly provable mathematics”) is computably enumerable. Hence, if there is any hope of a mechanistic model, the “product” of the imagined machine must be a set of (potentially) higher arithmetical
complexity.

Now, this train of thought leads directly to the aforementioned systems of Jeroslow’s design:

An experimental logic, is just a decidable ternary relation $H(t, p, \varphi)$ between natural numbers, with the intended interpretation “at time $t$, $p$ is recognized as (the code for) a proof of (the formula with the code) $\varphi$”. The set of theorems, $\text{Th}(H)$, is taken to be the $\Pi^0_2$-set of recurring formulas:

$$\text{Rec}_H(\varphi) \iff \forall t \exists s > t \exists p H(s, p, \varphi)$$

Here, we (like Jeroslow) just look at cases where every theorem is “decided in the limit”, as it were. An experimental logic is called convergent if every recurring formula is stable, i.e., belongs to the $\Sigma^0_2$-set:

$$\text{Stbl}_H(\varphi) \iff \exists p \exists t \forall s > t H(s, p, \varphi)$$

(Note, in relation to the picture of axiomatic theory development above, that this does not place any restrictions on the “update mechanism” – apart from it being mechanic to ensure the decidability of $H$.)

And so it seems as if Hazen has a good case here: It seems quite legitimate to view a $\Delta^0_2$-set as “the product of a machine”, and the experimental model adheres to a semi-euclidian picture of mathematical activity. Furthermore, Gödel’s second incompleteness theorem can only survive in a relativized form, as is shown by an easy example in Jeroslow’s paper, so self-consistency can be “provable” (in the generalized $\Pi^0_2$ sense). This looks very much like how we wanted to handle Lucas’ argument above. Are we happy then? No, not really.

This paper explains why the experimental logic setting is not sufficiently different from the (extensionally correct but) intensionally incorrect consistency statements of Feferman – though there are in fact some interesting differences. There is no real sense in which an experimental logic modelling mathematical theorizing can know that it is consistent. Furthermore, if you do think that Lucas’ argument has some force, a relativized Lucas argument (e.g., based on 2-consistency) can be devised for the experimental case.

Time permitting, some further properties of $\Delta^0_2$-sets, closely related to these concerns, will be touched upon.

References


