Research Plan

Joseph Tooby-Smith

The past

I started my academic career with a PhD and a postdoc in theoretical physics. During my time as a physicist, my speciality involved applying techniques from pure mathematics to solve problems in physics.

One area of mathematics that I used extensively was category theory. Category theory is seen by some as a unifying language for mathematics. My work showed how it can be used as a tool for interpretation and as a way to derive new results in physics. This is very similar to its use in computer science and specifically functional programming languages. I went on to specialise in an area of higher category theory called higher topos theory. Using this I derived new results related to an area of physics of current interest, called generalised symmetries. I'm proud of this work because it demonstrates an explicit use of higher category theory in deriving new results.

My work in higher topos theory resulted in me learning about homotopy type theory (the two have a subtle relationship). Homotopy type theory, itself, then resulted in me learning about theorem proving languages such as coq and Lean. I came to understand the potential benefit that such programs could have to the physical sciences. Thus, I made it my goal to promote the use of interactive theorem provers in the physical sciences, and overcome the hurdles prohibiting their use. This work, which has been self-directed and independent of faculty-involvement, is what I would like to be known for in the future.

In pursuit of this goal, my research turned from physics to computer science. My research turned into the use of underlying structures in functional programming, and applying them directly or indirectly to make interactive theorem provers easier for physical scientists. Solidifying this transition for my second postdoc, I joined a computer science department at Reykjavik University working with Tarmo Uustalu.

My most important work in this area to date, is the development of a library called HepLean (for which there is a corresponding preprint in the latter stages of review). This is similar to the project Mathlib and aims to formalise results from high energy physics in Lean 4. The project HepLean makes clear the benefits the physical scientist can gain from using a well-structured library written in an interactive theorem prover. Let me briefly mention some of these:

- It stores information from the area in a linear fashion, making look-up easier.
- It opens the door for the automatic derivation of new results in the physical sciences using AI and other automated tactics for theorem proving.
- It allows the community to automatically review papers and results for mathematicalcorrectness. Mathematical correctness is not something usually reviewed for in some areas of physics.
- It opens the door for new pedagogical methods inboth computer science and the physical sciences by for example using Lean games.

Current and future work in interactive theorem proving

Let me turn to current and future work. As mentioned above, part of my goal is to overcome the hurdles prohibiting the use of interactive theorem provers in the physical sciences. One such hurdle is notation. Physicists often rely heavily on notational conventions, where in some cases information is left implicit. One such notational convention is index notation for tensors. Part of my current work is about the formalisation of index notation in Lean 4. There is a balancing act here as the physicist should be able to use index notation in Lean as they would on pen and paper without much regard to any 'hidden details'.

The hidden details, however, have to be formal enough that Lean accepts them, and general enough to cover the different types of index notation used by physicists. To solve this problem, I am turning (perhaps unsurprisingly), to category theory, and specifically the theory of modular and coloured operads.

There are other hurdles similar to the notation one just discussed, that prohibit the use of interactive theorem provers in the physical sciences. One important hurdle is the learning curve that physicists must undergo to learn Lean and the current lack of reward they get from doing so. As a way to reduce the learning curve I plan to utilise informal definitions and lemmas in Lean. These are 'english' written results which can be later formalised by experts in Lean (or even better by AI), but are much easier for the physicist to write. There is a crude implementation of this already in HepLean. In addition I plan to learn how AI can be used to formalise the physical sciences and so increasing part of the reward for physicists. There is already a lot of work on using AI to formalise mathematics (e.g., DeepMind's work on Math Olympiad problems), which may not transition to the physical sciences smoothly, due to differing conventions between the communities.

Future work in category theory and computer science

As part of my future work, I plan to explore the intersection of category theory and computer science. I am specifically interested in the role higher category theory and higher topos theory can play in computer science. A key reference for this work will be Jacob Lurie's book on higher algebra, the material of which I am familiar with from my use of category theory in physics. In Lurie's book, higher generalisations of notions appearing in computer science appear, such as monoids and monads. Part of my future research will be investigating how these play a role, if any, in computer science.

Plan for student involvement

HepLean offers numerous opportunities for involving students in research. I plan to develop three lists of undergraduate-level and graduate-level projects around HepLean:

- Functional programming projects: An example of such a project would be the handling of lists in Lean to efficiently undertake computations needed for index notation (a notation used by physicists to deal with tensors). Other examples will involve meta-programming in Lean to make the user-experience easier.

- AI in Physics and Mathematics: These projects will explore auto-formalization of theorems in physics (converting human-written results into Lean proofs) and the inverse process, "auto-informalization." While these techniques have been explored in mathematics, they remain largely unexplored in the physical sciences, as dicussed above.

- Interdisciplinary Theorem Proving: These projects will involve proving physics theorems using Lean. Many such problems require minimal prerequisites in physics once the theorem is stated. For instance, formalizing properties of the two-Higgs doublet model potential could be an excellent project for students. Physicists are interested in its properties, such as its minima, whether it is bounded or not etc.

Each of these project lists will include homework-style tasks that can be completed in a few hours and more detailed thesis-level projects. The breadth of HepLean makes it relatively easy to generate a diverse range of projects.