In this paper I discuss the place of some computational formal methods in doing history of philosophy. Specifically, I describe how to apply interactive theorem provers in textual interpretation and argument reconstruction, to the benefit of both researchers and their broader scholarly community. More concretely, such applications involve formalizing key notions in the argument or text in a manner that the program can read and understand, compiling the file and fixing any runtime errors, and then identifying the philosophical results of such program executions. This last step usually involves producing a human-readable writeup of what was done, problems found, solutions implemented, and lessons learned, perhaps with some excerpted code from the program.

All of this can occur alongside and as a supplement to research produced using more traditional, informal, or non-computational methods in history of philosophy, that is, interactive theorem provers are not a replacement or substitute for critically thinking about a text as historians of philosophy have done for thousands of years, nor are they a replacement for longstanding methods in history of philosophy. Rather, deploying interactive theorem provers can complement and support the usual sort of activity, especially by automating for readers much of the mental labor of verifying arguments and spotting informal and formal fallacies. Perhaps most importantly, utilizing interactive theorem provers can spare others the labor of rewriting formalized arguments again once it has been done once because the source code from argument reconstructions in interactive theorem provers can be open-source. Thus, such code can be downloaded, modified, retooled, and fit to new purposes.

For any historian of philosophy, particularly if one is unfamiliar with interactive theorem provers, the natural question to ask at this stage is, 'Why would I do all of that?' What I described sounds like much more work for not terribly much payoff. After all, arguments of past philosophers can be reconstructed and even formalized on paper, as they have been for over a century, without the need to translate them into some interactive theorem prover's system. Such translations might even negatively effect the work: whichever formal system is used within some interactive theorem prover may have a distorting effect on the past philosopher's argument. So such applications of interactive proof assistants appear at first blush to involve much work with little or no gain, and, as a historian of philosophy sensitive to issues of translation is well-positioned to notice, could even be a substantial step backwards.

In this paper I address these concerns. My view is that interactive theorem provers have already been, and stand to continue being, useful to historians of philosophy. So much may seem obvious after reviewing some research in history of philosophy that leverages interactive theorem provers, which I do below. The novelty in my argument here is to indicate the untapped potential of interactive theorem provers to historians of philosophy. Interactive theorem provers cannot do philosophy for us, or, to make a more modest claim, nothing in my argument hinges on claiming that they can. But the manners in which interactive theorem provers can assist research in history of philosopher are about as plenitudinous as the ways in which computer verification and sharing code assists software development. That is what I argue for here. If this conclusion is true, then interactive theorem provers can be very useful tools indeed.

This paper builds on work by other philosophers in a similar vein, especially those applying interactive theorem provers in philosophy. There have been at multiple applications of the contemporary metaphysical and epistemological notions in philosophy. For example: Fitelson and Zalta (2007) have done axiomatic metaphysics in the interactive theorem prover

Prover9. Benzmüller et al. (2015) have formalized various modal systems and the relations between them in the interactive theorem prover *Isabelle/HOL*. Novak (2015) has used the computer proof-assistant MetaPRL to formalize certain epistemological notions and then used that formalization in MetaPRL to analyze well-known puzzles like the Surprise Examination Paradox. Blumson (2021) has axiomatized classical mereology in *Isabelle/HOL*.

Additionally, interactive theorem provers have been applied to the texts of past figures, including philosophers. For example: Fleuriot (2001) has formalized arguments in Isaac Newton's *Principia Mathematica* using interactive theorem provers. Lokhorst (2011) has formalized Mally's deontic logic and meta-ethical principles in the interactive theorem prover *Prover9*. Alama et al. (2015) have formalized (an interpretation of) Leibniz's theory of concepts in *Prover9*. Benzmüller and Paleo (2015) and Fuenmayor and Benzmüller (2017) have formalized multiple readings of Gödel's ontological argument for the existence of God in *Isabelle/HOL*. Building on the informal work in (Smith, 2020), Koutsoukou-Argyraki (2019) has formalized in *Isabelle* some of Aristotle's proofs and meta-theoretical results concerning his syllogistic.

Citing all these developments, Kirchner et al. (2019, §4) have defended the "benefit from interdisciplinary studies in which computational techniques are applied" and shown some use for interactive theorem provers in metaphysics. Fuenmayor and Benzmüller (2018) discuss the use of interactive theorem provers in formalizing natural language arguments and describe their approach as "computational hemeneutics." As yet, though, philosophers have not considered the general applicability of interactive theorem provers in doing history of philosophy, especially by reference to the scholarly activities of historians of philosophy and to the specific issues raised in applying formal methods, including computational ones like interactive theorem provers, in doing history of philosophy. This lacunae exists in the literature despite the fact that answers to some significant methodological issues are implicitly assumed in some applications of interactive theorem provers just noted, especially in the formalizations of Leibniz's theory of concepts and Gödel's argument for the existence of God. Hence, there is a real need for the present essay.

I organize the paper as follows. First I briefly describe what interactive theorem provers are (§2). The purpose of doing that will be to show how these programs can be used in the philosophical historian's practice of formalizing arguments. Those already familiar with interactive theorem provers might skip this section, referring back to specific details as needed. Next I discuss the metaphilosophical issues raised by formalizing arguments in doing history of philosophy (§3). There I argue that what is commonly called rational reconstruction of arguments can benefit from formalization using interactive proof assistants, and further, that such argument formalization can serve as a helpful complement to the other kinds of investigation undertaken by historians of philosophy. Then I discuss some examples of applying interactive theorem provers in history of philosophy (§4). Considering these applications will support my claim in §5 that formalizing arguments using interactive theorem provers can benefit the practice of doing history of philosophy. Finally I tie all of this discussion together to offer a prospective view of what interactive theorem provers can assist historians of philosophy in doing (§6). To give away the ending, computationally verifying argument reconstructions using such programs offers definite benefits to philosophers working in history of philosophy. Thus interactive theorem provers can be a useful tool to an important activity, rational reconstruction, in doing history of philosophy.

References

Jesse Alama, Paul E. Oppenheimer, and Edward N. Zalta. Automating leibniz's theory of concepts. In *International Conference on Automated Deduction*, pages 73–97. Springer, 2015.

Christoph Benzmüller and Bruno Woltzenlogel Paleo. Interacting with Modal Logics in the Coq Proof Assistant. In L. D. Beklemishev and D. V. Musatov, editors, *Computer Science – Theory and Applications*, volume CSR 2015 of *Lecture Notes in Computer Science, Vol. 9139*, pages 398–411, Switzerland, 2015. Springer. doi: 10.1007/978-3-319-20297-6.

Christoph Benzmüller, Maximilian Claus, and Nik Sultana. Systematic verification of the modal logic cube in isabelle/hol. *arXiv preprint arXiv:1507.08717*, 2015.

Ben Blumson. Mereology. *Archive of Formal Proofs*, March 2021. ISSN 2150-914x. `https://isa-afp.org/entries/Mereology.html`, Formal proof development.

Branden Fitelson and Edward N. Zalta. Steps toward a computational metaphysics. *Journal of Philosophical Logic*, 36(2):227–247, 2007.

Jacques Fleuriot. *A Combination of Geometry Theorem Proving and Nonstandard Analysis with Application to Newton's* Principia. Distinguished Dissertations. Springer-Verlag, 2001. ISBN 978-1-85233-466-6. doi: 10.1007/978-0-85729-329-9.

David Fuenmayor and Christoph Benzmüller. Automating emendations of the ontological argument in intensional higher-order modal logic. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, pages 114–127. Springer, 2017.

David Fuenmayor and Christoph Benzmüller. Computational hermeneutics: An integrated approach for the logical analysis of natural-language arguments. In *International Conference on Logic and Argumentation*, pages 187–207. Springer, 2018.

Daniel Kirchner, Christoph Benzmüller, and Edward N. Zalta. Computer science and metaphysics: A cross-fertilization. *Open Philosophy*, 2(1):230–251, 2019.

Angeliki Koutsoukou-Argyraki. Aristotle's assertoric syllogistic. *Archive of Formal Proofs*, October 2019. ISSN 2150-914x. `https://isa-afp.org/entries/Aristotles_Assertoric_Syllogistic.html`, Formal proof development.

Gert-Jan C. Lokhorst. Computational meta-ethics. *Minds and machines*, 21(2):261–274, 2011.

Natalia Novak. Practical Extraction of Evidence Terms from Common-Knowledge Reasoning. *Electronic Notes in Theoretical Computer Science*, 312(24):143–160, April 2015. doi: https://doi.org/10.1016/j.entcs.2015.04.009.

Robin Smith. Aristotle's Logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2020 edition, 2020.