

REED COLLEGE MATHEMATICS DEPARTMENT

THESIS PROPOSAL FORM

NAME:	Drew Blount	STUDENT ID:	E38027
[I deleted the instructions to save space -Drew]			
Turn the form in to Kim Kadas (L-317) BEFORE noon on Wednesday, September 10th, 2014 by 11:30 a.m. either in person or via email			
1. In what general area or areas of mathematics would you like to write your thesis?			
Machine learning—that child of computer science that perhaps only recently realized that its father is statistics			
2. If you have specific topics in mind, list them here, with some brief elaboration about the nature or scope of your proposed project. Include the names of professors with whom you have discussed these topics.			
<p>I would like to implement the “EGO Algorithm” put forth in:</p> <p>Donald R. Jones, Matthias Schonlau, and William J. Welch. “Efficient Global Optimization of Expensive Black-Box Functions.” <i>Journal of Global Optimization</i> 13, no. 4 (1998): 455–92.</p> <p>As I've discussed with Jim (who has agreed to advise this thesis), I'd build a modular piece of software---whose modularity involves an intricate abstract type system--to implement the algorithm, and use that framework to apply the algorithm to several function domains. I'll of course investigate and interrogate the algorithm theoretically. This theoretical perspective will inform my analysis of the algorithm's effectiveness in each domain.</p> <p>I have discussed the EGO algorithm at length with Mark Bedau in the phil. dept., who is my mentor at the Reed Artificial Life Lab, and who has worked with the algorithm in the past. Mark has suggested a certain function domain that I am particularly fascinated by, which could be the crux of my thesis: predicting emergent chemical properties of RNA strands from their base-pair representation {CCGAAGU...}. This is a classic “expensive black-box function” in the sense of the paper title above, and Mark has access to a data set of length-20 RNA strands which would enable application of the EGO algorithm. This domain is a tiny model version of the immensely complex genotype-phenotype mapping of biology, and so “decoding” the mapping between RNA and chemical properties would be a very significant result (according to Mark, a positive result would be “easily publishable”). Yet the RNA-mapping task seems tractable by the EGO algorithm, evidenced by its similarity to previous domains successfully “learned”, such as that described in Caschera & Bedau et al. “Automated Discovery of Novel Drug Formulations Using Predictive Iterated High Throughput Experimentation.” doi:10.1371/journal.pone.0008546.</p> <p>Because the algorithm works only from data, such an application would attempt to solve biochemical questions without knowing any biochemistry---the analysis is entirely within the numerical realm of statistics, computer science, and their child machine learning.</p> <p>Once I've built a modular piece of software for applying EGO to a “black-box function” module which takes inputs and returns outputs, it will be trivial to apply EGO to the above domain as well as any other. I plan on finding several more domains of interest throughout the year, and Jim, Mark and I are all interested in how EGO performs in different settings.</p>			

REED COLLEGE MATHEMATICS DEPARTMENT

THESIS PROPOSAL FORM

3. List three potential advisers. Clearly indicate your degree of preference, if any.
1) Jim (has agreed to advise this thesis) 2) Adam 3) Albyn