

Statement of Purpose  
Ian Drosos

I am applying to UC San Diego to seek a PhD in Cognitive Science. My main research interests are human-computer interaction, artificial intelligence, and machine learning. I am currently studying Computer Science at North Carolina State University (NCSU) as a master's student doing research under Dr. Christopher Parnin. I plan to continue with a career in research after obtaining my PhD.

I started my Computer Science education at Southern Polytechnic State University (SPSU) where I received my BS. My senior capstone project involved providing a new time-tracking system to SPSU's computer science department. This system took the existing paper version of graduate assistant time reporting and moved it online. The system allowed graduate assistants to fill out a report on hours spent on certain work activities. This report was then forwarded to the assistant's professor for approval. Once approved, the data from the report was archived for statistical generation by the administrator. The system generated graphs (using the Google Chart API) for the administrator to view.

After graduating I began work as a full-stack software engineer for Verizon. I worked on several systems during my four years there: ECPD (Enterprise Corporate Profile Database) which creates, modifies, and stores every profile of companies that Verizon sells to, EPOD (Enterprise Purchase Order Database) which creates and executes purchase orders for companies purchasing from Verizon, and CMS (Contract Management System) which handles the creation, modification, and execution of all enterprise contracts between Verizon and other companies. While working for ECPD I created an API that allowed the creation and modification of user roles to enable and control access to our systems. I was individually recognized for my contributions to a project for CMS that saved Verizon over 12 million dollars a year. I also was the designer and programmer of new functionality that allowed users to generate and execute non-disclosure agreements between Verizon and their customers. While my time at Verizon gave me valuable software engineering skills it did not allow me to pursue a career in research, so I applied to graduate schools so I could gain research skills and experience.

I joined NCSU as a master's student in Computer Science in 2015. Eager to start getting research experience, I started an independent study under Dr. Parnin to begin the project HappyFace which has become my master's thesis. The goal of HappyFace is to discover frustrating experiences felt during the programming process by collecting measures of frustration from learners. Learning-obstacles can interrupt the learning process and cause frustration in a learner if proper feedback mechanisms are not in place. For example, a learner met with a frustrating obstacle can begin to feel "stuck" when their efforts to fix an issue fall flat, causing them to wait until office hours to discuss their issue with the professor rather than continue with their learning experience. In addition, for large-scale learning systems traditional mechanisms such as peer feedback and tutor instruction do not scale well. Further in these large-scale systems, there is limited context about the problems being solved and the issues being faced by the learner. Discovering these obstacles is crucial in removing barriers to learning. Doing so in scale provides valuable information to MOOCs and other large-scale learning platforms, allowing them to provide fixes for these frustrating experiences. For this we performed a large-scale collection of code snippets from learners using PythonTutor, and collected a frustration rating through a light-weight feedback mechanism. We created abstract syntax trees (ASTs) from code snippets and processed the code for language features and other stylometric features like types of whitespace, code complexity, and the standard deviation of line length. These scores were then run through randomized logistic regression for feature selection to identify potentially frustrating learning experiences. We found that syntax errors, use of niche language features, and understanding complex code most correlated with frustration scores. We then extended HappyFace to allow for participants to select additional reasons

for frustration, annotating their experience. Using our code metrics, we then could predict the reason for frustration selected by the learner. Our findings are currently being compiled into a spring submission for VL/HCC 2017. For my thesis, I am extending the HappyFace system to provide interventions to the automatically detected obstacles to lower learner frustration and help the learner become unstuck.

For CSC 720 (Artificial Intelligence 2) I developed a similar feature extraction and prediction system. The goal of this project was to extract stylometric features (whitespace usage, code complexity, keyword usage, etc.) from JavaScript files and use those features to predict the author's gender. Using the GitHub archive of repositories matched with a confirmed gender table generated by Dr. Parnin's lab I gathered JavaScript files to parse into ASTs using Esprima. Feature data was gathered from the traversal of these ASTs and predictions were made using various prediction algorithms. Using cross-validation the most successful prediction algorithm was logistic regression with an average precision of 78% for predicting the code author's gender.

I am applying to the UC San Diego Cognitive Science department because of their promotion of research in learning and interaction, which are two of my interests. I also admire the interdisciplinary nature of the program. The opportunity to work with such a breadth of expertise is exciting. One professor I would like to work with is Dr. Philip Guo, whom I have had the privilege of working with during the HappyFace project. Working with him has been critical to the design and implementation of HappyFace and he has inspired my research direction of creating better tools for learners. Dr. Guo's recent works involving PythonTutor: Codepourri, Codeopticon, and Codechella have created advances in and have worked to lower barriers to programming education, a goal shared by the HappyFace project. I also am interested in Dr. Steven Dow's work in using crowdsourcing to elicit ideas and innovations as my work has also used a crowd to discover frustrating experiences in programming. Dr. Scott Klemmer's work with improving massive online classes also speaks to me as tools that empower people to learn is a research area of interest for me. I believe I will be able to provide meaningful contributions to research at the Cognitive Science department of UC San Diego if accepted.

After four years as a software developer and two years as a graduate student I would like to continue researching and building systems that help people. UC San Diego's focus on research will allow me to continue to grow as a researcher and contribute to the Cognitive Science field. I believe I am a good fit as my research interests closely align with your program. I look forward to continuing my research career as a PhD student at UC San Diego.