Past experience has shown my ability to perform research and to lead. Through research I solved real problems by using and expanding machine learning. I learned both recent advances and how much more there is to explore. Through leadership, I addressed practical problems and helped others. I want to pursue advanced study because it will enable me to combine research and leadership to have greater impact.

At Google, I created a feature that helps users find interesting and relevant content on Picasa Web Albums, a photo hosting site with millions of users. The difficult and most exciting part was extracting meaning from text and images provided by users. Text was often short and confusing to natural language processing algorithms such as Bayesian networks. However, I was able to adapt existing algorithms to the text available and produce a useful model. Yet text is only a proxy for image content. Ideally, an algorithm could process an image to identify important features, such as prominent objects. It would enable blind people to hear a description of the image, retailers to research product use, and users to find relevant but unlabeled images. This is beyond current image processing algorithms, which must be carefully trained to recognize a small class of objects. Waiting for others to work on my problems does not satisfy me. The Graduate Research Fellowship Program will enable me to work on bigger and more general problems that have wide-ranging impact.

To share Google's research outside the company, I am working on a cluster computing course. The course offers students access to thousands of processors, teaches tools to use them, and sponsors student projects. In January, I will lecture at MIT on problems encountered during my research and how to solve them with distributed systems. In addition, I am contributing course material, sample problems, and mentorship. While the subject is computer science, we are encouraging other students to come as large computing clusters are useful and necessary in many fields. Women and minorities, both underrepresented in the computer science department, are being encouraged to attend as we will favor projects that help them advance in computer science or consider it as a major. Google has already announced plans to expand the course, and I hope to bring it to a wider audience.

I acquired my knowledge of natural language processing while working as an intern in Banglore, India. Infosys, India's second largest software outsourcing provider, had me develop a way to automatically identify modules from disorganized source code. Since many others had tried structural information such as the call graph with mixed results, we took a novel approach that extracts topics from identifiers, such as variable and function names. I chose to use Latent Dirichlet Analysis (LDA) to process the names and experimented on our data. It worked well enough that we submitted a paper to the Software Mining Conference and Mark Steyvers, one of the researchers behind LDA, thanked me for the paper. The algorithms out there and opportunities for improvement got me excited about the field, so much so that I decided to continue in the field at Google. But it is the process of developing new models, mixing mathematical and practical concerns, and finding how to train and use them that interests me most. Advanced study provides me with the expertise necessary to pursue this interest and to lead development in natural language processing.

A Summer Undergraduate Research Fellowship with Netlab, the Caltech networking lab that periodically breaks Internet speed records, introduced me to machine learning and to the skills needed to understand natural language processing. My project was first described as doing something interesting with machine learning algorithms that model points in high dimensional spaces while working very independently. From there I looked at the methods

available and found many that learn manifolds in various ways, but few of them that provided some measure of confidence in the fit or the predictions it makes. So I chose one method, kernel principal component analysis, and created an error model for it. This has a variety of applications, such as finding outlier points, estimating confidence in a point's classification, and evaluating the fitness of parameters. Professor Low hired me to continue work on applications beyond the summer and I applied it to network security by finding outlier packets. It was this summer that led me to continue work in machine learning. This application is a continuation of that desire.

The very same summer I was with Netlab, dorm construction left hundreds of research students without Internet access. So I led the construction and operation of a student-run network. It started when I setup a wireless link and routers to cross the street. After that, I setup switches, led efforts to string cable, and convinced people to lend hardware. Eventually, every section of the dorm had access. The IT and housing departments were initially skeptical, with concerns ranging from viruses to tripping hazards. I had daily meetings with administrators as the network grew. With the IT department, I negotiated a temporary fiber uplink to replace the wireless link. The Vice President of Student Affairs toured the dorm with me and supported the network's presence. That summer, I practiced research and leadership, two skills that I will combine through advanced study.

Another chance for leadership came when I often heard complaints about Caltech's course registration system. As a student representative to two Caltech committees on computing, it was my responsibility to communicate these problems. Demonstrating a complete rewrite seemed like the best approach and so I led a ten week four person class project to do just that. We spoke with the registrar to understand their needs and obtain a term's worth of course offerings. I redesigned the database, made a completely new interface, wrote session management, and handled special-case courses. Other team members contributed core selection logic, a calendar display, conflict resolution, user management, and design documentation. During and after the project, I worked with and presented to administrators to gain adoption of our ideas. Finding real problems, demonstrating solutions, and advocating them are key parts of science. I led this project and expect to continue leading as a scientist.

I aspire to lead a computer science research group with a focus on machine learning. Currently natural language processing is my specialty, but the field changes rapidly and there will always be new topics to explore. I want to think about and solve hard problems with real impact while helping others do the same. Whether in academia or corporate research, I expect to publish results, implement solutions, and to educate the public about them. By funding advanced study, the NSF provides me with the knowledge, skills, and experience necessary to get there.

Experience with machine learning, research, and leadership has made me thirst for more. Graduate school is the place for me to work on interesting problems and contribute results to the research community and the public. At the same time, I will combine and enhance my research skills to reach my career goals. The NSF Graduate Fellowship Research Program will provide me with the freedom to work on difficult and experimental problems, to share results, and to continue contributing after the fellowship is complete.