Research Statement

My major research interests lie in machine learning and decision-making, and in the interaction of research in these fields with agent-based simulation and modeling. My work lies at the intersection of computer science and economics, and has deep interactions with the behavioral sciences and decision theory. From an algorithmic and computational standpoint I am mostly concerned with challenging problems in which agents must learn and make decisions over time in environments in which theoretically optimal behavior is hard or impossible to achieve. Examples of such problems include difficult variants of the classical bandit problem and emerging problems in algorithmic trading. From the systems science perspective, I use algorithms developed for these problems to model systems, like financial markets and hiring processes, in more complex and realistic ways. I also have more general interests across machine learning and information retrieval.

Computational Finance: My interests in computational finance center on market microstructure, the study of how the rules that govern trading affect price formation. The nature of trading in traditional stock markets is changing. Information is now more freely available, and a large proportion of trades are initiated by automated trading systems. There has also been a rapid increase in new kinds of electronic marketplaces which must decide on what kinds of structures and rules to use. Because of these developments, the study of microstructure has assumed increased practical importance.

Many aspects of exchanges and price dynamics have received a lot of attention recently from an algorithmic perspective. My research has focused on two related problems. First, how does one design effective trading algorithms given some assumptions about market structure? Second, what are the implications of the presence of agents using these algorithms on price dynamics?

I have investigated these problems in the context of algorithms for *market-makers*, agents specially designated by exchanges to provide liquidity and orderly price transitions in markets [1, 2], and in the context of algorithms that can be used by traders with superior information to optimize their expected profits [3]. My work on market-making takes the first step in the academic literature towards bridging the gap between stylized theoretical models and real markets. The design of a market-making algorithm capable of setting real dollar-and-cent prices for the first time enables comparison between time series and distributional properties of prices and returns in standard theoretical models and in real markets.

In addition, the algorithmic aspects of these trading problems turn out to be interesting in and of themselves. In particular, there is a connection between the exploration-exploitation tradeoff as thought of in reinforcement learning and sequential optimization, and deciding how to set prices or what quantities to trade. Applying knowledge from reinforcement learning to trading problems is a fruitful research direction whose surface has barely been scratched. The results of studies along these lines will yield knowledge that is applicable broadly outside the particular domain of financial markets.

I am particularly excited to explore these problems in the context of electronic markets which may be populated largely or exclusively by automated trading agents. Such markets can range from prediction markets to electronic crossing networks to markets for computational and network resources. The properties of these markets are not well understood so far, and research will be able to contribute extensively to both the design of agents to trade in these markets, and the design of the markets themselves so that they can more efficiently achieve the outcomes they are designed for, the efficient allocation of resources.

Agent-Based Modeling and Simulation My interest in computational finance developed because financial markets are a particularly clean example of a complex multi-agent system, providing clear agent objectives and interaction rules. However, I have always been interested in more general agent-based modeling. Some of my earlier work that began during my time as an undergraduate focused on the role of social norms in producing desirable group outcomes [4, 5]. In graduate school, I became interested in problems of search and matching, which led to a paper that examined the exploration-exploitation tradeoffs of agents who must learn their preferences in matching markets [6]. This was the first paper I am aware of that did not assume either that matching is exogenous and random (the standard assumption in the literature on two-sided search in economics), or that agents have *a priori* knowledge of their own preferences (the standard assumption in the matching literature in game theory and theoretical computer science). Our results demonstrated that the particular mechanism used to match agents in any given period can have substantial effects on the eventual outcomes of repeated matching games.

I also became interested in one of the subproblems faced by agents in this model of two-sided matching with learning. This subproblem relates to the literature on search and so-called *secretary problems* in statistics and operations research. This interest led to a more analytical paper about a problem which often arises in the process of searching for a job or for a candidate to fill a position [7]. Typically, applicants do not know if they will receive an offer from any given firm with which they interview, and, conversely, firms do not know whether applicants will definitely take positions they are offered. We quantified the value of information in the following sense: how much better off is a decision-maker if she knows each time whether she receives an offer or not, compared to the case where she is only informed when offers actually appear? We showed that she can expect to receive very close to her optimal value even in the lower information case, as long as she knows the probability that any given offer will appear. However, her expected value can fall dramatically when she does not know the appearance probability *ex ante* but must infer it from data. This suggests that hiring and job-search mechanisms may not suffer from serious losses in efficiency or stability from participants hiding information about their decisions, unless agents are uncertain of their own attractiveness as employees or employers.

While this last example is not strictly on agent-based modeling, it does provide insight from a systems perspective by solving a single-agent problem explicitly. The methods developed in this paper can then be used to model agents in more complex systems. In general, I think the field of agent-based modeling can benefit from more complex agent models. Endowing agents with the best algorithms known for certain problems is a good model of research to counter arguments that agent-based modeling is ad-hoc. I have elaborated on this view of modeling, and its relation to research in bounded rationality, in a recently submitted paper [8]. I plan to continue investigating algorithms for agents that must learn and make decisions in increasingly complex environments, and then using these algorithms to model complex systems.

Machine Learning and Information Retrieval My interests in machine learning and information retrieval are broad. My first academic paper was on feature selection in a supervised learning setting [9], and I have maintained interests in all aspects of machine learning throughout graduate school. My time as a post-doctoral scholar at UCSD has allowed me to start thinking about practical problems in machine learning and information retrieval from a research perspective, and I have been excited to dive into a number of problems.

One of the projects I am working on is a scientific workflow for automating the process of updating specialized biomedical databases from rich sources of data such as the primary scientific literature and existing general databases [10]. Participating in a real case-study has introduced me to the divide between theoretical results, which tend to focus on tuning algorithms to improve performance on standard datasets, and practical problems, where framing the problem well becomes much more important than the particular choice of algorithm.

I have also become interested in network analysis. It turns out that there are interesting parallels between research on axiomatic ranking systems in the economic and decision-theoretic literatures, and algorithms like PageRank. My postdoctoral advisor Charles Elkan and I are attempting to exploit these parallels to understand information structures in different domains that can be represented as networks, including citation analysis and college sports. This work is at a preliminary stage, but we have recently received internal funding from the University of California to investigate some of these directions. I am excited about having a chance to investigate these domains, and I am sure that I will find synergies between this research and my work on agent-based systems that I can investigate in the future.

Conclusion While my research spans different subfields of machine learning and computer science, there are several themes that occur repeatedly. The first of these is a focus on learning behavior, especially the learning behavior of agents who must repeatedly interact with, and make decisions in, complex environments. The second theme is the importance of understanding information structures, ranging from modeling asymmetric information in markets to understanding the information present in the network of pairings in college football games. The third theme is understanding the behaviors of systems as they emerge from interactions between the agents who constitute the systems.

My goals in research are twofold. I want to provide scientific understanding of agents and systems, and I want to use this understanding to help us learn how to design systems (ranging from coordinated agent systems to electronic marketplaces) more efficiently by better predicting the impacts of structural changes.

While there are many research directions I am exploring, and many more that I am excited to explore, much of my inspiration comes from understanding the domains that my colleagues are interested in and applying my expertise to bring fresh perspective to the problems they think about. I hope to be in an environment where I continue to be stimulated by the ideas of my colleagues, and where my ideas in turn provide stimulation for them.

References

[1] Sanmay Das. A learning market-maker in the Glosten-Milgrom model. *Quantitative Finance*, 5(2):169–180, April 2005.

[2] Sanmay Das. The effects of market-making on price dynamics. Submitted for publication, 2006.

[3] Sanmay Das. Learning to trade with insider information. Technical Report Center for Biological and Computational Learning Paper 255 / Artificial Intelligence Lab Memo 2005-028, Massachusetts Institute of Technology, October 2005. Submitted for publication.

[4] Barbara J. Grosz, Sarit Kraus, David Sullivan, and Sanmay Das. The influence of social norms and social consciousness on intention reconciliation. *Artificial Intelligence (Special Issue on Multi-Agent Systems)*, 142(2):147–177, November 2002.

[5] Sanmay Das, Barbara Grosz, and Avi Pfeffer. Learning and decision-making for intention reconciliation. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 1121–1128, Bologna, Italy, July 2002.

[6] Sanmay Das and Emir Kamenica. Two-sided bandits and the dating market. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, pages 947–952, Edinburgh, UK, August 2005.

[7] Sanmay Das and John N. Tsitsiklis. When is it important to know you've been rejected? A search problem with probabilistic appearance of offers. Technical Report # 2699, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, May 2006. Submitted for publication.

[8] Sanmay Das. On agent-based modeling of complex systems: Learning and bounded rationality. Submitted for publication, 2006.

[9] Sanmay Das. Filters, wrappers, and a boosting-based hybrid for feature selection. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 74–81, Williamstown, MA, June 2001.

[10] Aditya Kumar Sehgal, Sanmay Das, Milton H. Saier, Jr., and Charles Elkan. Automatically updating a specialized protein database: Comparing literature-based and database-based approaches. Submitted for publication, 2006.