Research Statement

Summary I am interested in (1) the design of intelligent agents and systems, primarily guided by machine learning; (2) modeling and understanding collective dynamics that result from intelligent individual behavior; and (3) using this understanding to inform the design of venues where people and automated agents come together to interact. A central focus of my research is on understanding how information flows through systems, how it can be best used by intelligent agents, and how its presence, absence, or the form in which it is available impacts decisions at the individual and systemic levels. My work can be categorized into four broad themes.

1: Collective intelligence I am interested in both modeling and understanding the dynamics of collective intelligence, and in designing algorithms that allow us to use the power of collective wisdom to make better decisions. I have been working on the foundations of a rigorous theory of how information grows in novel social media like **Wikipedia and the blogosphere**, and on information aggregation and dissemination in prediction markets. In recent work, we have documented some remarkable regularities in the life cycles of average Wikipedia pages and blog posts [26, 27]. They exhibit a concave rise to an editing / commenting peak, followed by decay at a $1/t$ rate over time. We have proposed a simple model of information creation that matches the data well. This model is the first step in the development of a comprehensive model of information creation in such "collective wisdom processes." I am also actively pursuing related questions, including the dynamics of commits to open source software, and the development of algorithmic techniques for detecting attempts to manipulate public opinion through Wikipedia.

Prediction markets aggregate the beliefs of many agents with different information about an event (like "Barack Obama will win the presidential election") into a single number, a price. Prediction market prices have been shown to reflect the true probabilities of outcomes closely [50, 8]. Recently there have been broad calls to deploy markets for socially important tasks, like predicting the passage of climate change legislation [5]. Much of my work in this area has focused on algorithmic liquidity provision from a reinforcement learning perspective (described in more detail below). We have also recently started running experimental prediction markets with human participants, including the *Instructor Rating Markets* at RPI, which allow students to trade on the ratings their professors will receive. Many interesting incentive issues arise when market participants also affect the outcomes they are trading on (in this case by providing ratings). Our markets were successful: students participated actively in both trading and regularly rating their instructors, and prices were predictive of future ratings, thus providing *dynamic* feedback to instructors on the progress of their classes [13].

2: Reinforcement Learning I often work on reinforcement learning problems motivated by economics and markets, but the techniques are broadly applicable. For example, one of the major challenges in designing and deploying prediction markets is in generating liquidity in the market so that people are willing to trade [49]. Much of the work in this area focuses on designing liquidity providing market makers using the framework of scoring rules, which incentivize truthful reporting (at least myopically), but suffer from the problem that they are in general loss-making [32, 16, 39]. Others have considered market-making as a trading strategy, but without the market-maker being obliged to take the other side of every trade [45, 15]. My own work [21, 22, 25] frames the problem of market making from a reinforcement learning perspective within established models of asymmetric information. In this framework, the problem is essentially one of dynamic pricing, and is of interest for reinforcement learning because prices simultaneously serve two roles: they are both the "sensors," giving information about the distribution of valuations in a population (will someone buy or sell at this price?), as well as the profit-making mechanism. This sets up an exploration-exploitation dilemma.

Models that incorporate an optimizing, reinforcement learning market maker can predict interesting characteristics of market behavior. For example, we show that a profit maximizing monopolistic market maker may actually provide more liquidity than a zero-profit one in times of market uncertainty, because she is willing to take short-term losses in order to learn more quickly [25]. This model provides support for anecdotal claims from the New York Stock Exchange about the value of their "specialist" model (equivalent to a monopolistic market maker) in times of high market uncertainty.

Beyond the phenomenological predictions of the model, the theoretical development has allowed us to design a practical algorithm for market making that we are testing in both simulation and experiments with human subjects. One new focus will be to test the algorithms in simulation settings with intelligent trading agents, inspired by various trading agent contest scenarios [36, 48]. Our results so far indicate that this algorithm has the potential to offer lower average losses than the *de facto* standard market making algorithm based on the logarithmic market scoring rule, while also maintaining lower spreads and providing more liquidity [11, 13]. Variants of the algorithm have been used in both subsequent academic work by several different groups (e.g. [10, 46]), as well as in the private sector in finance.

One of the core algorithmic elements in the market making algorithm is the idea of using moment-matching approximations to maintain tractable belief states that agents can efficiently represent and perform inference on. This idea turns out to be broadly powerful beyond just market making. We have used this technique for the setting of posted-price "digital goods" auctions, where a seller is attempting to sell items with no marginal cost of production (for example, music or movie downloads) [17]. Prior to our work, most research on this problem was either completely theoretical (distribution-agnostic regret bounds for simple algorithms) [35, 9], or restricted to relatively simple models of uncertainty [43, 19]. We are among the first to design practical algorithms for these problems and evaluate them in complex environments. We have also used the idea of moment matching approximations to tackle a classic problem in online inference: "noisy bisection," where a learner has to track a target and can place a sensor: she then only sees a *thresholded signal* (whether or not the target is "above" or "below" the sensor), and this signal may be noisy [12, 33, 47]. We show that our technique allows us to perform asymptotically almost as well as we could with actual signals, instead of thresholded ones [14].

I have also worked on reinforcement learning for improving human-robot interaction [38] and reinforcement learning in nonstationary multi-agent systems [24].

3: Search, matching, and multi-agent systems Market design is a naturally interdisciplinary field, because of the importance of both algorithmic ideas and a need to understand the allocation of scarce resources. I have worked on markets where costly search plays a significant role [37], as well as on markets that match pairs or groups of agents [42]. The design of matching markets has been used for important social purposes, like matching applicants to jobs (most famously, medical school graduates to their first residencies [40]), and matching kidney donors to recipients

[41, 1, 6]. But the matching literature has typically assumed that agents know their preferences in advance. We explored the consequences of agents having to learn their preferences sequentially through interactions with each other, in the context of a "dating game" [24]. This was among the first papers to look at matching when preferences have to be learned (it has since been followed by several papers in the economics literature that examine matching with unknown preferences), and revealed the critical importance of the mechanism used for matching on the long-term stability of outcomes. Another paper looks at the effects of learning in search processes with exploding offers, like academic hiring markets, and shows that, for an applicant, the information gleaned from rejection signals is much higher when she is unaware of her own "worth" or attractiveness in the market, by enabling her to estimate this more quickly [29].

Another application of search theory is to e-commerce. Current models of e-commerce marketplaces are almost exclusively two-sided, considering direct interactions between buyers and sellers [31]. In reality, the presence of intermediaries can have significant effects on buyer and seller strategies, especially through the incentives they are willing to offer to the intermediaries [7, 34]. In a collaboration with David Sarne of Bar-Ilan University (funded by the US-Israel Binational Science Foundation), we are studying the concept of *expert-mediated search*, best illustrated by an example. Consider a consumer looking for a used car on a large Internet marketplace. She sees noisy signals of the true value of any car she looks at the advertisement for, but she can disambiguate this signal by paying for the services of an expert (for example, getting a Carfax report, or taking the car to a mechanic for an inspection). We are examining how the presence of the expert changes the consumer's search process, and the role that can be played by a market designer or regulator in increasing social welfare in such markets [18].

Recently I have been worked on understanding social welfare in matching markets [3, 4]. This work analyzes matching from the perspective of cardinal utility models, where the quantity of utility derived from a matching is important, not just the preference ordering [2]. I am also interested in models of multi-agent systems outside of economic markets: I have worked on the importance of social norms in multi-agent teams [23, 30], as well as on models of disease spread that leverage real data on transportation networks [51].

4: Supervised learning I am interested in both the theory and applications of supervised learning. I have worked on new algorithms for feature selection in supervised learning [20], and on training set selection when the data consists of only labeled examples of one class, and all other examples are unlabeled: we designed an approach to solving this problem using minimal effort from an expert human labeler [28]. The latter research forms the foundation for an automated pipeline for triage of proteins that may be added to a curated, specialized biomedical database [44].

I am excited about two new projects related to supervised learning. First, as a visiting scholar at the US Treasury's Office of the Comptroller of the Currency, I am applying machine learning methods to predict and manage systemic risk across large financial institutions, leveraging a unique dataset with detailed time series tracking of the behavior of individual credit card account holders. Second, we have just received a new NSF award to work on humanitarian logistics in collaboration with experts on disaster relief: one of our tasks will be to use existing data (both hand-collected and scraped from the web and social media) to predict the flow of donations to disaster sites; we hope to build an understanding of the determinants of donation behavior that can then be used to influence donor behavior in socially useful directions.

References

[1] D.J. Abraham, A. Blum, and T. Sandholm. Clearing algorithms for barter exchange markets: enabling nationwide kidney exchanges. In *Proc. ACM EC*, 2007.

[2] Elliot Anshelevich and Sanmay Das. Matching, cardinal utility, and social welfare. *ACM SIGECom Exchanges*, 9(1), 2010.

[3] Elliot Anshelevich, Sanmay Das, and Yonatan Naamad. Anarchy, stability, and utopia: Creating better matchings. In *Proceedings of the Second Symposium on Algorithmic Game Theory (SAGT)*, pages 159–170, 2009.

[4] Elliot Anshelevich, Sanmay Das, and Yonatan Naamad. Anarchy, stability, and utopia: Creating better matchings. *Autonomous Agents and Multi-Agent Systems*, 2011. Forthcoming.

[5] K.J. Arrow, S. Sunder, R. Forsythe, R.E. Litan, M. Gorham, E. Zitzewitz, R.W. Hahn, R. Hanson, D. Kahneman, J.O. Ledyard, et al. Statement on prediction markets. AEI-Brookings Joint Center Related Publication No. 07-11, 2007.

[6] P. Awasthi and T. Sandholm. Online stochastic optimization in the large: application to kidney exchange. In *Proc. IJCAI*, pages 405–411, 2009.

[7] Y. Bakos. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science*, 42:1676–92, 1997.

[8] J. Berg, R. Forsythe, F. Nelson, and T. Rietz. Results from a dozen years of election futures markets research. *Handbook of Experimental Economic Results*, pages 486–515, 2001.

[9] A. Blum, V. Kumar, A. Rudra, and F. Wu. Online learning in online auctions. *Theoretical Computer Science*, 324(2-3):137–146, 2004.

[10] K. Boer, U. Kaymak, and J. Spiering. From discrete-time models to continuous-time, asynchronous modeling of financial markets. *Computational Intelligence*, 23(2):142–161, 2007.

[11] Aseem Brahma, Sanmay Das, and Malik Magdon-Ismail. Comparing prediction market structures, with an application to market making. *Arxiv preprint arXiv:1009.1446*, 2010.

[12] M.V. Burnashev and K.S. Zigangirov. An interval estimation problem for controlled observations. *Problemy Peredachi Informatsii*, pages 51–61, 1974. Originally in Russian.

[13] Mithun Chakraborty, Sanmay Das, Allen Lavoie, Malik Magdon-Ismail, and Yonatan Naamad. Instructor rating markets. In *Proceedings of the Second Conference on Auctions, Market Mechanisms, and Their Applications*, 2011. Abstract appears.

[14] Mithun Chakraborty, Sanmay Das, and Malik Magdon-Ismail. Near-optimal target learning with stochastic binary signals. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, 2011.

[15] T. Chakraborty and M. Kearns. Market making and mean reversion. In *Proc. ACM Conf. on Elec. Commerce*, 2011.

[16] Y. Chen and D. Pennock. A utility framework for bounded-loss market makers. In *Proc. UAI*, pages 49–56, 2007.

[17] Meenal Chhabra and Sanmay Das. Learning the demand curve in posted-price digital goods auctions. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 63–70, 2011. **Nominated for the Best Student Paper Award (one of three nominees)**.

[18] Meenal Chhabra, Sanmay Das, and David Sarne. Expert-mediated search. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 415–422, 2011.

[19] V. Conitzer and N. Garera. Learning algorithms for online principal-agent problems (and selling goods online). In *Proceedings of the 23rd international conference on Machine learning*, pages 209–216. ACM, 2006.

[20] 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.

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

[22] Sanmay Das. The effects of market-making on price dynamics. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 887–894, Estoril, Portugal, May 2008.

[23] 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.

[24] 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.

[25] Sanmay Das and Malik Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. In *Advances in Neural Information Processing Systems (NIPS)*, pages 361–368, 2008.

[26] Sanmay Das and Malik Magdon-Ismail. Collective wisdom: Information growth in wikis and blogs. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 231–240, 2010.

[27] Sanmay Das and Malik Magdon-Ismail. A model for information growth in collective wisdom processes. *ACM Transactions on Knowledge Discovery from Data*, 2011. Forthcoming.

[28] Sanmay Das, Milton H. Saier, Jr., and Charles Elkan. Finding transport proteins in a general protein database. In *Proceedings of the Eleventh European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 54–66, Warsaw, Poland, September 2007.

[29] Sanmay Das and John N. Tsitsiklis. When is it important to know you've been rejected? a search problem with probabilistic appearance of offers. *Journal of Economic Behavior and Organization*, 74:104–122, 2010.

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

[31] R. Guttman, A. Moukas, and P. Maes. Agent-mediated electronic commerce: A survey. *Knowledge Engineering Review*, 13(2):147–159, June 1998.

[32] R. Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, 2003.

[33] B. Jedynak, P. L. Frazier, and R. Sznitman. Twenty questions with noise: Bayes optimal policies for entropy loss. *Journal of Applied Probability*, 2011. To appear.

[34] J. Kephart and A. Greenwald. Shopbot economics. *JAAMAS*, 5(3):255–287, 2002.

[35] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for on-line posted-price auctions. In *Proc. FOCS*, 2003.

[36] J.K. MacKie-Mason and M.P. Wellman. Automated markets and trading agents. *Handbook of computational economics*, 2:1381–1431, 2006.

[37] J. McMillan and M. Rothschild. Search. In R. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications*, pages 905–927. North-Holland, 1994.

[38] Eric Meisner, Sanmay Das, Volkan Isler, Jeffrey Trinkle, Selma Sabanovic, and Linnda R. Caporael. Predictive state representations for grounding human-robot communication. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 178–185, 2010.

[39] A. Othman, T. Sandholm, D. Pennock, and D. Reeves. A practical liquidity-sensitive automated market maker. In *Proc. EC*, pages 377–386, 2010.

[40] A. E. Roth and Elliott Peranson. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748–780, 1999.

[41] A.E. Roth, T. Sönmez, and M.U. Ünver. Kidney Exchange. *Quarterly Journal of Economics*, 119(2):457–488, 2004.

[42] Alvin E. Roth and Marilda Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monograph Series. Cambridge University Press, Cambridge, UK, 1990.

[43] I. Segal. Optimal pricing mechanisms with unknown demand. *American Economic Review*, 93(3):509–529, 2003.

[44] Aditya Sehgal, Sanmay Das, Keith Noto, Milton H. Saier, Jr., and Charles Elkan. Identifying relevant data for a biological database: Handcrafted rules versus machine learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(3):851–857, 2011.

[45] A. Sherstov and P. Stone. Three automated stock-trading agents: A comparative study. *Agent-Mediated Electronic Commerce VI. Theories for and Engineering of Distributed Mechanisms and Systems*, pages 173–187, 2005.

[46] S. Stathel, J. Finzen, C. Riedl, and N. May. Service Innovation in Business Value Networks. In *Proceedings of the XVIII International RESER Conference*, 2008.

[47] R. Waeber, P.I. Frazier, and S.G. Henderson. A Bayesian Approach to Stochastic Root Finding. In *Proc. Winter Simulation Conf.* IEEE, 2011.

[48] M.P. Wellman, A. Greenwald, and P. Stone. *Autonomous bidding agents: Strategies and lessons from the trading agent competition*. The MIT Press, 2007.

[49] J. Wolfers and E. Zitzewitz. Five open questions about prediction markets. Stanford GSB Working Paper, 2004.

[50] Justin Wolfers and Eric Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.

[51] Teruhiko Yoneyama, Sanmay Das, and Mukkai Krishnamoorthy. A Hybrid Model for Disease Spread and an Application to the SARS Pandemic. *Journal of Artificial Societies and Social Simulation*, 2011. Forthcoming.