

## Research Statement

R. Teal Witter

I love research: the challenge of a difficult problem, the connection between two seemingly different ideas, the surprise when my intuition fails, and the deep insight when an idea finally clicks.

As an algorithmic generalist, I leverage mathematical insights to tackle computational problems. While my research spans a diverse set of domains, I am often guided by the following questions: How can we *design* better—more accurate, robust, efficient, and simple—algorithms for the problem at hand? How can we *analyze* these algorithms to mathematically understand their properties and limitations? By considering these questions across disciplines, I have built an extensive practical and theoretical toolkit. I particularly enjoy applying this toolkit in under-resourced domains where algorithmic solutions have the potential for significant social impact.

My research reflects the many wonderful people who have taught and inspired me. I have co-authored 10 peer-reviewed papers (and four preprints) with researchers from across seven academic institutions, learning from and collaborating with brilliant people on questions that fascinate us. I have advised three undergraduate research projects, guiding four students through their first research experiences. I have worked closely with one nonprofit organization, leveraging my expertise to support the expansion of their work. Looking forward, I am eager to establish new collaborations with colleagues from diverse backgrounds, mentor student projects that drive meaningful change, and focus my research on problems with significant social impact. By doing so, I aim to continue producing innovative research that benefits the academic community and society at large.

### Machine Learning for Social Good

The primary focus of my research is applying machine learning to projects with the potential for social good. I have worked to improve methods for explainable AI, measure the impact of nonprofits from natural experiments, and simplify fairness algorithms.

In the age of big data, explaining machine learning predictions is crucial to deploying AI in high-stakes domains. Inspired by game theory, Shapley values are a widely used method for attributing predictions to  $n$  data features. However, exactly computing Shapley values requires  $\Omega(2^n)$  time, so approximation algorithms are often necessary. In recent work, I proposed a single-line modification to one of the most popular methods for computing Shapley values, Kernel SHAP. Inspired by an elegant mathematical connection to linear regression, the modification improves the algorithm's accuracy, enabling a simple implementation to outperform the extensively optimized official version. While there is no theoretical guarantee on the performance of Kernel SHAP, I show that the modified Kernel SHAP is provably accurate: With high probability, the modified Kernel SHAP algorithm returns a solution which is within a  $(1 + \epsilon)$  multiplicative factor of the true Shapley values with only  $\tilde{O}(n/\epsilon^2)$  time. The result proves that Shapley values can be effectively approximated by (a modified version of) Kernel SHAP in almost linear time.

Measuring the impact of a treatment from a natural experiment is a crucial problem across the social sciences. In recent work, I collaborated with Reach Out and Read Colorado (RORCO), an early childhood literacy nonprofit that provides free books and a 'prescription for reading' to children at pediatric check-ups. Measuring the impact of RORCO is particularly challenging because treatments are correlated with outcomes by design: they focus their efforts on the students with the lowest literacy levels who need their help the most. Treatment effect estimation is a well studied problem but, when I applied 20+ algorithms to the RORCO data, I found the algorithms were disturbingly inconsistent. In order to evaluate the algorithms,

I built an extensive benchmark with synthetic outcomes designed in consultation with domain experts.<sup>1</sup> The benchmark indicates that the class of doubly robust algorithms gives the best performance. To understand why, I exactly computed the finite variance of doubly robust estimators with a training-testing split and, motivated by the exact variance, proposed a new doubly robust algorithm with a novel loss function. Together, the benchmark and theoretical analysis lay a comprehensive foundation for the practical and theoretical development of treatment effect estimators in the natural experiment setting. Already, RORCO has leveraged my findings to boost their fundraising and expand the reach of their work.

Ensuring fair machine learning predictions is crucial for high-stakes applications. In the fairness domain, my research has focused on designing simple and effective algorithms. In a recent project, I developed a comprehensive benchmark for fairness algorithms that incorporate uncertainty into their predictions.<sup>2</sup> The benchmark revealed that incorporating uncertainty into predictions can reduce group differences, as measured by statistical parity, without any fairness intervention. In another fairness project, I explored a complicated fairness metric called counterfactual fairness under a mild assumption [7]. Counterfactual fairness attributes outcomes to unobservable features (e.g., intelligence and work ethic) and characterizes predictions as ‘fair’ if they only depend on these features. When unobservable features are statistically equivalent between demographic groups, as they are in social settings, I showed that counterfactual fairness is mathematically equivalent to the far simpler notion of statistical parity. Taken together, my research suggests that simple algorithms can achieve more fair outcomes than complicated approaches.

## Optimization with Social Impact

My research leverages optimization techniques to tackle complex problems with social implications. I have applied deep learning to urban planning, utilized complexity theory to formulate abstract social good problems, and used linear algebraic insights to model opinion dynamics.

Living in New York City, I have seen firsthand the cost of traffic congestion. One innovative approach is the ‘open streets’ initiative, where streets are closed to cars and opened to people. However, choosing which streets to open is a challenging optimization problem that requires careful consideration of traffic patterns, weather, infrastructure, and vehicle accidents. In my work, I designed a deep reinforcement learning algorithm that leverages graph neural networks (capturing spatial relationships) and recurrent components (capturing temporal relationships) to minimize congestion and collisions [10].<sup>3</sup> This data-driven approach outperforms existing methods and offers a more equitable solution compared to the existing application-based method that favors well-resourced neighborhoods.

Restless multi-armed bandits (RMABs) offer a powerful framework for solving resource-constrained problems with social impact. However, the traditional maximization formulation can be inappropriate for settings with a reward requirement such as reintroducing endangered species in wildlife conservation, reducing energy use in sustainability, and supporting patient recovery in healthcare. In recent work, I propose a new RMAB formulation that minimizes cost subject to a reward requirement. Like the traditional maximization problem, I show that the minimization problem is PSPACE-hard and, instead of provable algorithms, must be solved with heuristics. While similar in some ways, there are RMAB instances where standard heuristics exactly solve the maximization problem but give the worst possible performance on the minimization problem. I propose novel heuristics specifically designed for the minimization problem and demonstrate their potential for superior performance. My work suggests the importance of the minimization problem and the need for specially designed algorithms.

<sup>1</sup><https://github.com/rtealwitter/NaturalExperiments>

<sup>2</sup><https://github.com/rtealwitter/FairlyUncertain>

<sup>3</sup><https://github.com/rtealwitter/OpenStreets>

The formation of political opinions is an important component of a healthy democracy. Prior work has attempted to simulate the polarization we observe in politics with complicated models. In my work, I instead propose a relative measure of polarization that accounts for the perception of increased polarization [6]. Under this new measure, the most simple model of opinion dynamics provably reflects the polarization we observe. Surprisingly, my work explains political alignment across multiple, unrelated topics: If opinions are purely a product of a social network then people will only agree (or only disagree) on almost every topic. By taking a linear algebraic perspective, my work simplifies opinion dynamics, providing mathematical insights into opinion polarization.

## Discrete Optimization

Theoretical insights are important for understanding the advantages and limitations of algorithmic techniques. As a simple testbed, I have analyzed algorithms for problems related to Boolean function evaluation. On the classical side, I explored the performance of adaptive and evolutionary algorithms relative to their counterparts. On the quantum side, I have leveraged algorithms for Boolean function evaluation to develop efficient algorithms for graph theory problems.

Understanding the trade-offs inherent in algorithm design is key to effective deployment. An algorithm that is adaptive can update its behavior as it discovers more about the input. In my work, I analyzed the gap between the best adaptive and non-adaptive algorithms for Boolean functions [3]. My results show that the adaptivity gap can vary from  $\Theta(\log n)$  for simple functions like DNF formulas to  $\tilde{\Theta}(n)$  for more expressive functions like read-once formulas. The results suggest that adaptivity is generally beneficial unless the function is simple or adaptivity is prohibitively expensive. In addition to adaptivity, I have also explored the power of local search algorithms with any-time guarantees. In my work, I proved that a local search algorithm that outputs a diverse set of solutions has a constant approximation factor that approaches that of the best known greedy algorithm [4]. The results suggest that local search algorithms can be a powerful complement to greedy algorithms especially when solutions should be diverse and robust.

Since building quantum computers is so challenging, designing efficient theoretical algorithms is important. In my quantum computing research, I have worked extensively with a beautiful semi-definite program (SDP) that gives provably-optimal algorithms for Boolean function evaluation. I have used the SDP to design nearly optimal algorithms for graph theory problems like bipartite detection and maximum matching [2, 5]. My work relates one problem to another through fine-grained reductions, giving performance guarantees in terms of problem-specific parameters. However, designing algorithms from analytical solutions to the SDP is notoriously difficult. I showed that numerical solutions to the SDP can still give query-optimal quantum algorithm even when the SDP is only solved approximately [1]. The results open the door to numerically-derived quantum algorithms that are nearly query-optimal.<sup>4</sup>

Beyond the research described in the three arms above, I have applied mathematical ideas to board games: I have analyzed the computational hardness backgammon [8] and explored Ticket to Ride from a graph theoretic perspective, evaluated heuristic strategies that leverage effective resistance [9].

## Future Work

I am excited to leverage the skills and insights I have developed in my research to tackle new problems. As a professor, I will focus my research agenda on impactful problems with both theoretical and empirical angles. This approach will allow students to participate in research projects even without a strong theoretical background. As an example of my research agenda, I will describe two projects that I am excited to pursue.

I am eager to extend my work on estimator design to other areas with social impact. The first project

---

<sup>4</sup><https://github.com/rtealwiter/QuantumQueryOptimizer>

is in the context of wildlife conservation. Mark and recapture—where animals are observed in subsequent observations—is a common technique in ecology to estimate the size of a wildlife population. However, estimators often rely on strong assumptions that are difficult to verify. I plan to develop estimators that incorporate deep learning and offer guarantees even when common assumptions are violated.

And another one.

## Conclusion

Throughout my research, I have been driven to tackle complex problems, exploring connections between seemingly disparate areas and developing novel algorithmic solutions. My work has spanned a diverse set of domains, from machine learning to quantum computing. I have been privileged to collaborate with wonderful researchers, domain experts, nonprofits, and students, yielding impactful contributions at the intersection of theory and practice. Looking forward, I aim to leverage my research experience to address impactful social problems, foster interdisciplinary collaborations and mentor students in an inclusive and supportive environment.

## References

*\*As is the custom in theoretical computer science, authors are listed in alphabetical order unless otherwise specified with an asterisk.*

- [1] M. Czekanski, S. Kimmel, and R. T. Witter. Robust and space-efficient dual adversary quantum query algorithms. In *European Symposium on Algorithms*, 2023.
- [2] K. DeLorenzo, S. Kimmel, and R. T. Witter. Applications of the quantum algorithm for st-connectivity. In *Conference on the Theory of Quantum Computation, Communication and Cryptography*, 2019.
- [3] L. Hellerstein, D. Kletenik, N. Liu, and R. T. Witter. Adaptivity gaps for the stochastic boolean function evaluation problem. In *Workshop on Approximation and Online Algorithms*, 2022.
- [4] L. Hellerstein, T. Lidbetter, and R. T. Witter. A local search algorithm for the min-sum submodular cover problem. In *International Symposium on Algorithms and Computation*, 2022.
- [5] S. Kimmel and R. T. Witter. A query-efficient quantum algorithm for maximum matching on general graphs. In *Algorithms and Data Structures Symposium*, pages 543–555, 2021.
- [6] C. Musco, I. Ramesh, J. Ugander, and R. T. Witter. How to quantify polarization in models of opinion dynamics. In *International Workshop on Mining and Learning with Graphs*, 2022.
- [7] L. Rosenblatt and R. T. Witter. Counterfactual fairness is basically demographic parity. In *AAAI Conference on Artificial Intelligence*, 2023.
- [8] R. T. Witter. Backgammon is hard. In *International Conference on Combinatorial Optimization and Applications*, 2021.
- [9] R. T. Witter\* and A. Lyford. Applications of graph theory and probability in the board game ticket to ride. In *International Conference on the Foundations of Digital Games*, 2020.

- [10] R. T. Witter\* and L. Rosenblatt. I open at the close: A deep reinforcement learning evaluation of open streets initiatives. In *AAAI Conference on Artificial Intelligence*, 2024.