

## 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 am particularly excited to apply this toolkit in under-resourced domains where algorithmic solutions have the potential for significant social impact.

My academic research is really the story of many wonderful people. 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 in 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 impact of a treatment is a crucial problem in many social settings. In recent work, collaborate with RORCO to measure impact of their work. Particularly challenging because treatments are correlated with outcomes by design: direct literacy help to the students who need it most. Well studied problem and evaluate effect according to 20+ estimators, but observe differing results. Create benchmark using synthetic outcomes designed in consultation with domain experts. Observe doubly robust estimators give best performance, including on five other causal inference datasets. To avoid complicated statistical dependence, theoretically investigate doubly robust estimators with training-testing split. Exactly characterize finite variance in this setting and propose algorithm with novel loss function inspired by variance characterization. Already, impact to RORCO fundraising, help with expansion of their program.

Making predictions fairly is crucial to positive impact. In recent work, incorporate unavoidable uncertainty in fair way. How to make prediction when uncertain? Answered in haphazard way by prior work. Unify setting with extensive benchmark for both classification and regression tasks. Find that a simple algorithm gives best uncertainty predictions in classification setting. Propose uncertainty-aware statistical parity definition In regression setting, find that incorporating uncertainty can decrease statistical parity without any explicit intervention. Before, simplify complicated definition of fairness under a mild assumption on groups. Counterfactual fairness gets attributes source of performance to ‘latent’ attributes and only uses latent attributes for predictions. We show mathematically that if latent attributes are statistically the same, as they are between demographic groups, counterfactual fairness is actually equivalent to the far simpler definition of statistical parity. Results show how simple algorithms meet and even exceed more complicated approaches for achieving fairness.

A fundamental problem for nonprofit organizations is to measure the impact of their work. Ideally, nonprofits would use a longitudinal randomized controlled trial to measure the effect of their ‘treatment’, but this is often infeasible due to cost and ethical concerns. In my work, I developed and theoretically analyzed an algorithm that estimates treatment effect from observational data even when treatment is assigned in a non-random way. The algorithm uses a regression adjustment with a novel loss function that is robust to non-random treatment assignments. I collaborated with Reach Out and Read Colorado (RORCO), a nonprofit that provides free books and a ‘prescription for reading’ to children at pediatric check-ups, to tune the algorithm to address their setting and evaluate its performance on their data.<sup>1</sup> Already, RORCO has used my results to bolster their fundraising and expand their work.

I have also worked on several projects to improve the fairness of machine learning algorithms. When making predictions in real life, algorithms necessarily struggle to capture the underlying randomness of the world. Understanding this uncertainty and accounting for it in algorithms is important for making fair decisions. In my work, I built an extensive benchmark to evaluate methods for incorporating uncertainty into fairness-aware algorithms. Leveraging the benchmark, I developed a novel algorithm for regression problems incorporating uncertainty that can reduce unfairness without any explicit fairness intervention.<sup>2</sup> In another fairness project, I theoretically analyzed the relationship between two popular fairness definitions, counterfactual fairness and demographic parity. While counterfactual fairness is difficult to measure, I showed that it is equivalent to demographic parity under a common assumption in social settings [7]. Since demographic parity is easier to measure and design algorithms to achieve, my results offer a simpler method to achieve and measure counterfactual fairness.

## Optimization with Social Impact

Living in New York City, I have seen firsthand the negative impact of traffic congestion. An increasingly popular approach to improve urban environments is to close streets to cars and open them to people, a practice known as ‘open streets’. Interestingly, because of a phenomenon known as Braess’s paradox, open streets can actually reduce congestion on the remaining streets by eliminating bottlenecks. However, choosing which streets to open is a challenging optimization problem that requires evaluating traffic patterns, weather, infrastructure, and likely accidents. In my work, I designed a deep reinforcement learning algorithm to select streets to open that minimize congestion and collisions [10]. The algorithm uses graph neural networks to incorporate the complex spatial relationships of urban infrastructure and a recurrent component to capture the temporal dynamics of traffic and weather. The algorithm was evaluated on real data from the New York City Department of Transportation and showed a reduction in congestion and collisions compared to the current method of opening streets. In contrast to the current application-based approach

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

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

that biases the benefit towards well-resourced neighborhoods, the model offers an objective and data-based approach for selecting open streets.<sup>3</sup>

Restless multi-armed bandits are a classic problem in sequential decision-making that exemplify the trade-off between exploration and exploitation. Traditionally, the goal is to maximize reward obtained from activating a sequence of arms subject to a budget constraint. In my work, I extended the restless bandit formulation to include a minimization objective, allowing for more flexibility than the standard budget constraint. I found a reduction to polynomial-space Turing machines, showing the minimization problem is PSPACE-hard, just like the maximization problem. However, in experiments and theoretical results, I demonstrated that minimization problem is fundamentally different and requires new algorithms and insights. My work is a first step towards understanding the minimization problem.

Opinion dynamics [6]

## Discrete Optimization

Theoretical understanding is an important component of designing algorithms that are both efficient and accurate. Leveraging the relative simplicity of sequential decision-making problems, my research theoretically explored the properties and limitations of different algorithmic techniques. In particular, I explored the relationships between optimization problems, the power of adaptivity in decision-making, and the effectiveness of local search algorithms.

In addition to exploring the relationship between optimization problems, I have worked on testing the limits of algorithmic techniques in the context of Boolean function evaluation. Given a Boolean function on  $n$  bits, the goal is to evaluate the function on an unknown input with as few queries as possible. An adaptive algorithm can query the function and use the results to decide which input to query next, while a non-adaptive algorithm must decide all queries in advance. In my work, I analyzed the gap between the best adaptive and non-adaptive algorithms for several classes of 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 for sequential decision-making problems. 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 or robust.

Quantum computers have the potential to revolutionize computing by harnessing the quantum mechanical properties of superposition and entanglement. However, because building quantum computers at scale is so challenging, it is important to theoretically analyze how useful quantum computers will be. My research has focused on designing and analyzing quantum algorithms for graph theory problems and developing new techniques for numerically deriving quantum algorithms.

My research began with a focus on designing quantum algorithms for graph theory problems. In my work, I designed and analyzed efficient quantum algorithms for problems related to graph connectivity like cycle and bipartite detection [2]. I applied several reductions to reduce cycle and bipartite detection to  $st$ -connectivity. Because of the fine-grained nature of the reductions, the guaranteed complexity of the quantum algorithms incorporated problem-specific parameters. In a follow-up work, I adapted an efficient classical algorithm for maximum matching to design a quantum algorithm with the best known quantum query complexity for maximum matching on general graphs [5].

---

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

I then turned to developing new techniques for numerically deriving quantum algorithms. The algorithms I previously designed originate from a semi-definite program (SDP) whose solutions correspond to query-optimal quantum algorithms. However, analytically solving the SDP is difficult because of its complexity. In my work, 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 areas above, I have worked on several projects informed by my collaborators and the questions that interest me. I analyzed the computational hardness of the board game backgammon [8]. I explored the board game Ticket to Ride from a graph theoretic perspective and 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 with social impact. As a professor, I will focus my research agenda on impactful problems with both theoretical and empirical angles. In this way, students can 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 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.

The second project is in the context of explainable AI. A popular approach to explaining the predictions of a machine learning model is to assign ‘credit’ to the features of the input, quantities known as Shapley values. However, exactly computing Shapley values can require an exponential number of computations with respect to the number of features. I plan to leverage a beautiful theoretical connection to linear regression and develop an estimator that provably approximates Shapley values with only a linear number of computations.

## **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.

---

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

## References

- [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.

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