I use tools from theoretical machine learning and algorithmic game theory to understand the relationship between (predictive) algorithm design and algorithmic decision-making. Algorithmic predictions supplement human and algorithmic decision-making in a variety of domains: predicted risk of disease leads to flagged scans for additional review from medical professionals, predicted hyperparameters change physics models for simulator design in engineering, and predicted market trends change investment strategies in quantitative finance. Such algorithmic predictions are often made by training a model which minimizes some *loss* function measuring error, and predictions are used to guide some downstream *decision* or recommendation. Approximately accurate algorithmic predictions can often lead to poor decision-making and recommendations if these algorithms are not able to make "smarter" errors when necessary.

For example, if, given a CT scan x, a patient's true risk of disease is $p^*(x) = 0.06$, then predictive models $h_1(x) = 0.04$ and $h_2(x) = 0.08$ are both "equally inaccurate" in terms of squared error. Suppose a hospital's policy is to run more extensive tests if a model yields prediction $h(x) > 0.05$, then h_1 leads to no tests being run, while h_2 recommends additional tests. In this case, the decision made with h_2 aligns with the decision recommended by the true probability $p^*(x)$, while h_1 does not. Designing losses with this decision-like structure allows the algorithm to distinguish between these cases, allowing the model to make predictive errors that less starkly affect decision-making, as demonstrated in Figure 1.

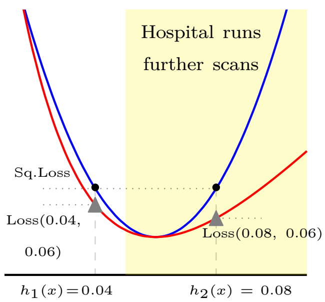

Figure 1: Example where the true risk is $p^*(x) = 0.06$, and models h_1 and h_2 have equal squared loss (blue). However, if the hospital uses predictions for decision-making, the stakes of these predictions change. My goal is to develop principled loss functions that incorporate decision-making processes into their structure, such as the loss in red.

Subtle challenges emerge when incorporating such structure into losses. Loss functions often need to balance several desiderata; aligning the loss with decision task is often at odds with designing losses that are computationally tractable to minimize (e.g., convex). Historically, convex surrogate losses have been constructed in an ad-hoc manner, and often do not align with the intended decision task, including surrogates for top-k prediction. Conversely, feasibility or equity concerns conceptually require algorithm designers to modify the optimized loss, but little work has characterized *how* various constraint formulations change decision-making. **My research agenda examines the bidirectional relationship between algorithm design and decision-making in machine learning and algorithmic economics**, and spans application settings ranging from resource allocation, information retrieval, high-stakes classification, and uncertainty quantification, among others.

Thread 1: **Decision-making to algorithm design.** Given a decision problem, design a "good" loss (part of the machine learning algorithm) that is statistically corresponds to the decision problem.

Thread 2: **Algorithm design to decision-making.** Given a fixed algorithm, examine how values embedded into the algorithm change decision-making.

My work in these threads has received recognition through an NSF Graduate Research Fellowship, an NSF Mathematical Sciences Postdoctoral Research Fellowship, and spotlight recognition at Neural Information Processing Systems (NeurIPS), awarded to the top 4% of submissions [2].

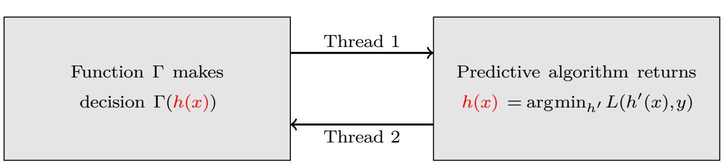

Figure 2: Thread 1 starts with a decision task and studies the design of predictive algorithms that correspond to said task, and Thread 2 examines the changes to decision-making induced by modifying a given algorithm.

Thread 1: Designing "good" algorithms for given decision tasks

In supervised machine learning, empirical risk minimization is the dominant paradigm for learning from data. Therein, one learns to predict by minimizing a *surrogate loss function* $L:\mathbb{R}^d\times\mathcal{Y}\to\mathbb{R}_+$ measuring the error of a prediction $u\in\mathbb{R}^d$ against the observed outcome $y\in\mathcal{Y}$, such as the hinge loss $L(u,y)=(1-uy)_+$, as a surrogate for binary classification. Ideally, minimizing this surrogate loss can be linked to making the correct decision for a given decision task or target loss. A decision task is formalized either by a function $\Gamma:\Delta_\mathcal{Y}\to\mathcal{R}$ mapping probability distributions (over outcomes) to decisions in $\mathcal{R}$ or by a target loss $\ell:\mathcal{R}\times\mathcal{Y}\to\mathbb{R}_+$ measuring the error of the report $r\in\mathcal{R}$ against the outcome $y\in\mathcal{Y}$. Examples of decision tasks include classification $\Gamma(p)=\text{argmax}_{y\in\mathcal{Y}}p_y$ (the most likely of outcomes in $\mathcal{Y}$), structured prediction tasks on graphs, and estimating risk measures such as variance or conditional value at risk, and target losses include 0-1 loss, least-squares regressions, Borda ranking losses, and cost-sensitive classifications. Since algorithms optimize surrogates $L:\mathbb{R}^d\times\mathcal{Y}\to\mathbb{R}_+$ as a proxy for targets defined over a possibly different set $\mathcal{R}$, a link function $\psi:\mathbb{R}^d\to\mathcal{R}$ is often necessary to reason about decisions given predictions. For example, the sign link $\psi:u\mapsto\textbf{sign}(u)$ often accompanies the hinge loss as a surrogate for binary classification. My work studies the design of surrogate losses and links that are **(a) *consistent* for a given target problem**: for all sequences of models mapping features in $\mathcal{X}$ to predictions $\{h_n:\mathcal{X}\to\mathbb{R}^d\}$ and all data distributions $D\in\Delta(\mathcal{X}\times\mathcal{Y})$, $\mathbb{E}_D L(h_n(X),Y)\to\inf_h\mathbb{E}_D L(h(X),Y)\implies\mathbb{E}_D\ell(\psi\circ h_n(X),Y)\to\inf_h\mathbb{E}_D\ell(\psi\circ h(X),Y)$. Roughly speaking, consistency means that minimizing the expected surrogate loss implies minimizing the expected target loss or making the best decision possible when linking surrogate predictions by ψ. Moreover, we want **(b) convex surrogates**, so that minimizing the surrogate loss function should be computationally tractable. In the sequel, I give a brief overview of my research findings on constructing convex and consistent surrogates in continuous estimation and discrete prediction settings; see Table 1 for the settings covered by each work.

	Target loss	Target decision
Discrete target	e.g., binary classification [3–5, 8, 9]	e.g., hierarchical classification [3, 5, 8, 9]
Continuous target	e.g., least-squares regression [2, 5]	e.g., variance estimation [2, 5]

Table 1: My work on constructing convex and consistent surrogate losses spans settings of both continuous and discrete estimation tasks, and applies when given a reference decision task or target loss function.

Discrete prediction tasks When the decision problem is discrete, such as top-k selection, rankings, and classification, constructing a convex surrogate loss function is less straightforward. Often, convex surrogates for discrete decision tasks are carefully and laboriously handcrafted, yet in practice, established surrogates are often inconsistent for the intended decision task. In [3], we develop the *embeddings framework* for designing consistent and convex surrogate losses for general discrete prediction tasks. Intuitively, given a discrete decision task, one can "embed" the possible decisions into the d-dimensional reals ($\mathbb{R}^d$) via an embedding function $\varphi:\mathcal{R}\to\mathbb{R}^d$ and "convexify" the space in between embedded reports. Using the convex conjugate, **we construct a convex (and piecewise linear) surrogate loss function that embeds a target loss** or decision task. Theorem 1 in turn implies this constructed surrogate loss is consistent for the target problem.

Theorem 1 ([3, Theorem 2]) *For any discrete target loss $\ell:\mathcal{R}\times\mathcal{Y}\to\mathbb{R}_+$, there exists a piecewise linear and convex surrogate $L:\mathbb{R}^d\times\mathcal{Y}\to\mathbb{R}_+$ and link $\psi:\mathbb{R}^d\to\mathcal{R}$ such that (L,ψ) is consistent with respect to ℓ.*

To prove Theorem 1, we characterize *all possible* consistent link functions $\psi:\mathbb{R}^d\to\mathcal{R}$, the construction of which is generally not intuitive. This link characterization enables one to analyze the consistency of *any* piecewise linear and convex surrogate loss and link pair, and has been used to prove the inconsistency of proposed surrogates for top-k classification [8], structured prediction [9], and the popular Weston-Watkins hinge for multiclass classification [12], highlighting conditions when these surrogates may fail. Moreover, for top-k classification, we construct a consistent surrogate and link, **resolving a conjecture about its (non)existence** [13].

Thread 2: Algorithm design's impact on decision-making

In consequential settings where predictions and decisions directly affect human livelihood, concerns surrounding the fairness and justice of decisions are salient. Machine learning (e.g., predictive algorithms) and algorithmic mechanism design (e.g., resource allocation algorithms) are becoming increasingly interconnected in algorithmic decision-making pertaining to human livelihood (e.g., [11]). Fairness in this joint context is not well-studied, as the approaches historically taken by the optimization and allocation communities differ substantially. In [6, 7], we enumerate opportunities and lessons for the machine learning and resource allocation communities to learn from each other, demonstrating these opportunities across five domains.

In predictive machine learning algorithms, constraints are often added to the objective function as a form of regularization function; in some sense, these regularized losses are new, generally unscrutinized losses, and the optimal decision they induce is unknown. **My work characterizes when and how these modified losses change the optimal decision**, exemplified by considering fairness regularizers [1]. In examining how fairness constraints change decision-making, we find that **only fairness constraints that value "accuracy as fairness" leave the optimal decision unchanged**; examples of such constraints in binary classification settings include calibration violations (the absolute difference) and bounded group loss constraints (reweighing the original loss). Unsurprisingly, the converse implies that many group fairness metrics such as equalized false positive rates and equality of opportunity change the optimal decision for certain outcome distributions; we characterize spaces of data distributions where decisions remain unchanged.

Future work and conclusion

My past and ongoing work explores the relationship between (tractable) prediction and decision-making, and I plan to further explore questions related to these threads.

T1: Decision → algorithm: trading off consistency and efficiency My previous work in the first thread has focused on the design of efficient surrogate loss functions that are simultaneously convex and consistent for *every* problem instance. If a decision maker has some prerequisite information about the problem, they may be willing to trade off consistency for efficiency. I am interested in understanding how we can improve efficiency if we only require consistency for practically relevant problem instances, rather than for *every* problem instance.

T1: Decision → algorithm: partnering with practitioners to design losses for spatial prediction While the embeddings framework is very general, I have shown firsthand its practical use for designing tractable algorithms for a variety of tasks. In particular, embeddings have tremendous untapped potential to improve predictions in structured, spatial domains. I plan to apply the embeddings framework to spatial prediction problems like risk prediction for wildfire risk (e.g., in Colorado) to inform decision-making around wildfire management such as the assignment of prescribed burns. I have been a member of the Mechanism Design for Social Good working group focusing on Conversations with Practitioners, through which I have learned about developing and sustaining collaborations that are mutually beneficial for practitioners and marginalized communities as well as researchers [10].

T2: Algorithm → decision: learning when to predict I am interested in understanding when the use of noisy predictive models is more beneficial to decision-making than a more reliable, but less granular, given statistic. In many settings, predictive models are both expensive to build and error-prone, and decision-makers have access to some summary statistic of the data that is both more accurate and less expensive to access, at the cost of granularity of information. In certain settings, predicting the more granular data can help the decision-maker distinguish between different instances that share the same value of the summary statistic. I plan to characterize when a predictive algorithm yields better decisions than algorithms relying solely on summary statistics, and vice versa.

Conclusion I am interested in understanding the relationship between prediction and algorithmic decision-making. My agenda is foundational, and contributes to understanding the design of predictive algorithms for various decision tasks such as information retrieval, building recommender systems, and risk prediction for allocating services such as cloud server storage or wildfire management services, among others.

References

[1] **Jessie Finocchiaro**. Using property elicitation to understand the impacts of fairness constraints. 2023. arXiv.

[2] **Jessie Finocchiaro** and Rafael Frongillo. Convex elicitation of continuous properties. In *Advances in Neural Information Processing Systems 31*, 2018.

[3] **Jessie Finocchiaro**, Rafael Frongillo, and Bo Waggoner. An embedding framework for consistent polyhedral surrogates. In *Advances in Neural Information Processing Systems 32*, 2019.

[4] **Jessie Finocchiaro**, Rafael Frongillo, and Bo Waggoner. Embedding dimension of polyhedral losses. In *The Conference on Learning Theory*, 2020.

[5] **Jessie Finocchiaro**, Rafael Frongillo, and Bo Waggoner. Unifying prediction bounds for consistent convex surrogates. In *Advances in Neural Information Processing Systems 34*, 2021.

[6] **Jessie Finocchiaro**, Edwin Lock, Faidra Monachou, and Manish Raghavan. Fairness and discrimination through the dual lens of mechanism design and machine learning. Economics and Computation 2021 Tutorial, July 2021. `https://sites.google.com/colorado.edu/fairness-tutorial-ec21/home`.

[7] **Jessie Finocchiaro**, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. Bridging machine learning and mechanism design towards algorithmic fairness. In *ACM Conference on Fairness, Accountability, and Transparency*, pages 489–503, 2021.

[8] **Jessie Finocchiaro**, Rafael Frongillo, Emma Goodwill, and Anish Thilagar. Consistent polyhedral surrogates for top-k classification and variants. In *Proceedings of the 39th International Conference on Machine Learning*, 2022.

[9] **Jessie Finocchiaro**, Rafael Frongillo, and Enrique B Nueve. The structured abstain problem and the lovász hinge. In *The Conference on Learning Theory*, 2022.

[10] **Jessie Finocchiaro**, Mackenzie Jorgensen, Yaren Bilge Kaya, Di Nguyen, Matthew Olckers, Mayra Russo, Kristen Scott, and Wendy Xu. Bridging research and practice through conversations with practitioners. 2023. In preparation.

[11] Jakob Schoeffer, Alexander Ritchie, Keziah Naggita, Faidra Monachou, **Jessie Finocchiaro**, and Marc Juarez. Online platforms and the fair exposure problem under homophily. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023. First five authors contributed equally.

[12] Yutong Wang and Clayton Scott. Weston-watkins hinge loss and ordered partitions. In *Advances in Neural Information Processing Systems*, 2020.

[13] Forest Yang and Sanmi Koyejo. On the consistency of top-k surrogate losses. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.