# Leveraging Human Input to Enable Robust AI Systems

Daniel S. Brown

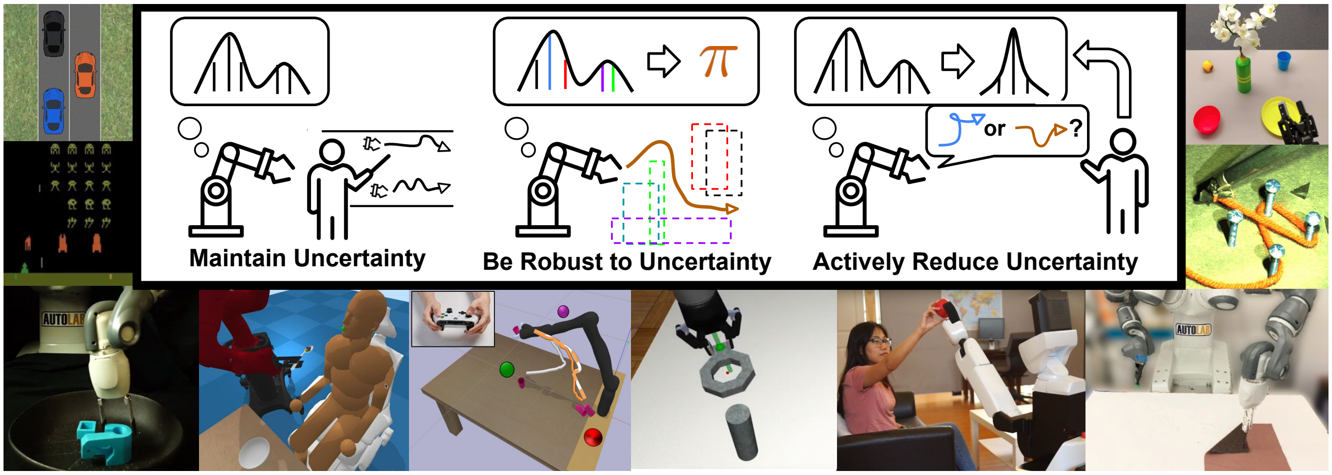

Figure 1: **My work seeks to directly and efficiently incorporate human input into both the theory and practice of robust machine learning.** I apply rigorous theory and state-of-the-art machine learning techniques to enable AI systems to maintain, be robust to, and actively reduce uncertainty over both the human's intent and the corresponding optimal policy. I evaluate my research across a range of applications, including autonomous driving, service robotics, and dexterous manipulation.

My goal is to develop **robust AI systems** that can be deployed in situations that are risk-sensitive such as self-driving cars, healthcare, and domestic service robots. A fundamental motivation behind my work is the idea that **what qualifies as *robust* is ultimately determined by the human user**: we want to be highly confident that the AI systems we use and interact with are going to do what we, as humans, actually want them to do. Much work on robust machine learning and control seeks to be resilient to, or completely remove the need for, human input. By contrast, I see human input as an extremely valuable and even critical component of robustness. However, making sense of human input requires dealing with sub-optimality, inconsistency, and ambiguity. To enable robust learning from human input, my work combines techniques from reinforcement learning, Bayesian inference, human-factors, deep learning, and operations research. My work bridges both theory and practice, with an emphasis on theory that is constructive and leads to significant and practical algorithmic improvements.

**Prior Research** My prior work considers robust value alignment in the context of **learned reward functions** that incentivize desired behavior. When teaching an AI system to perform a complex task, such as driving a car or folding a t-shirt, it is often easier to demonstrate the task, rather than write down an explicit reward function for the task. However, human demonstrations can be suboptimal [1, 2] and human preferences can be ambiguous [3, 4], motivating the need to maintain uncertainty over the human's true intent. My prior work developed the first algorithm for Bayesian reward function inference that scales to high-dimensional, visual imitation learning tasks [5], enabling AI systems to efficiently **maintain uncertainty** over a human's true intent. I also derived performance bounds for AI systems that learn from demonstrations that are four orders of magnitude more sample efficient than prior state-of-the-art bounds [6], allowing AI systems to have high-confidence in the correctness of behaviors learned from *practical* amounts of human input [7]. My work also addresses the problem of how an AI system should optimize its behavior to **be robust to uncertainty** over its objective function [8]. Leveraging mathematical tools from financial risk management, I developed a novel reinforcement learning algorithm that enables AI systems to hedge against risk by optimizing policies that are robust against multiple, possibly competing, reward function hypotheses [5, 3]. I also developed active learning algorithms that let AI systems **reduce uncertainty** [4, 9, 5] by querying for additional human input in ways that minimize supervisor burden [10, 11] while providing bounds on generalization error [12].

**Research Agenda** While my work so far has considered human input in the form of demonstrations and preference rankings, there is a wide range of potential human feedback types, explicit as well as implicit, that AI systems can and should leverage. In the future, I will develop theory and algorithms for actively learning from **multi-modal human input**, such as demonstrations, preferences, emergency stops, corrections, verbal instructions, and body language, to learn better models of human intent. Furthermore, while my prior work focuses on robustness in the form of probabilistic safety bounds, robustness is also tied to ideas from explanability and verification. In the future, I will develop tools for **interpretable and verifiable robustness**. I plan to use human input to assist AI systems to better convey their learned behaviors and assist humans in verifying that these systems are aligned with human values. Finally, learned reward functions are not the only aspect of the world to which AI systems need to be robust. In the future, I will develop techniques for human-in-the-loop robust machine learning that **generalize beyond reward function uncertainty**. I will develop safe and robust machine learning techniques that efficiently leverage human input to enable AI systems to actively reduce uncertainty over dynamics, object affordances, human rationality, safety constraints, and even the human's own mental model of the robot.

# 1 Prior Research

My prior work has primarily focused on the problem of robust **value alignment**: ensuring that robots and other AI systems do what we, as humans, actually want them to do. In particular, my research allows AI systems to learn behaviors that are robust to missing or mis-specified reward functions. Designing good reward functions is highly challenging and error prone, even for domain experts. Consider trying to write down a reward function that describes good driving behavior, how you like your bed made in the morning, or how to prepare a salad. While reward functions for these tasks are extremely difficult to write down, the desired behaviors can readily be demonstrated, even by non-roboticists. However, human data can often be extremely difficult to interpret, due to ambiguity and noise. Thus, it is critical that AI systems are robust to epistemic uncertainty over the human's true intent. Within the area of reward learning, I have focused on developing reward learning algorithms that (1) efficiently **maintain uncertainty** over human intent, (2) directly optimize behavior to be **robust to uncertainty** over human intent, and (3) actively query for additional human input to **reduce uncertainty** over human intent.

## 1.1 Maintaining Uncertainty: Efficient Bayesian Uncertainty using Human Input

Maintaining uncertainty is a critical first step towards robustness; however, prior research on reward learning either completely ignores uncertainty or uses computationally intractable algorithms which require solving hundreds or thousands of reinforcement learning problems to generate a posterior distribution over likely human reward functions. My work addresses this problem by developing the first **scalable Bayesian reward inference algorithm for visual imitation learning domains.** This research combines my earlier work on T-REX [1, 2], a state-of-the-art approach for learning reward functions from small numbers of ranked, suboptimal demonstrations, with self-supervised deep learning techniques, to learn a lower-dimensional latent state-representation where Bayesian reward inference becomes tractable. In comparison to prior Bayesian reward inference approaches, which would take *days to run*, my research enables Bayesian reward inference in only a *few minutes* by leveraging a small number of human pairwise preferences over trajectories. [5].

Leveraging efficient representations of reward function uncertainty, I used Bayesian counter-factual reasoning to develop a novel framework for computing high-confidence probabilistic safety bounds that are *four-orders of magnitude more sample efficient* than the prior state-of-the-art bounds [7, 6]. My research enables a critical component of robustness: the ability to provide practical high-confidence bounds on generalization performance, when learning from small numbers of human demonstrations. As a novel application of these high-confidence performance bounds, my research was the first to demonstrate a scalable and practical method to **automatically detect reward hacking or gaming behaviors**, cases where the robot's learned reward function results in a behavior that violates the human's true preferences.

## 1.2 Robust to Uncertainty: Optimizing Behavior that is Risk-Sensitive

Most prior work on imitation learning takes a risk-neutral reward learning approach that optimizes for expected performance, ignoring possibly catastrophic tail risk and leading to overly aggressive and unsafe policies [3]. By contrast, my work takes a risk-aware approach to path planning [13, 14] and policy optimization [5, 3] that explicitly accounts for uncertainty. My research enables AI systems to **hedge against uncertainty, rather than seeking to uniquely identify the human's reward function**.

Using duality theory, coherent risk metrics from mathematical finance, and policy gradient methods I developed a novel Bayesian robust policy optimization algorithm that optimizes an AI system's behavior such that it performs well under multiple, possibly competing, objectives inferred from human input [5]. On challenging high-dimensional control benchmarks, my approach optimizes policies that are significantly more robust than prior state-of-the-art approaches which struggle to effectively deal with small numbers of ambiguous human preferences [3]. By optimizing for multiple objectives simultaneously, my recent work allows AI systems to be robust to spurious correlations and also provides a practical solution for the increasingly important problem of multi-agent value alignment—optimizing decision policies that reflect and address the values and preferences of multiple groups or individuals.

### 1.3 Reducing Uncertainty: Actively Requesting for Additional Human Input

The previous two sections discuss prior work that uses human input to enable high-confidence performance bounds and robust policy optimization. However, what if an AI system has so much uncertainty over the human's true reward function that it cannot guarantee good performance with sufficiently high probability? My research addresses this problem via active learning: generating targeted queries for additional human input in states where the AI system believes it may have high generalization error. My work shows that risk-aware queries are more effective and efficient at reducing generalization error than more generic uncertainty reduction queries [12]. By combining active risk-aware queries with my prior work on high-confidence performance bounds, my research demonstrates, for the first time, that a **robot can know how many demonstrations it needs to learn a particular task** [7, 12]. I recently applied these techniques in a lifelong shared autonomy setting, where a robot assists a human when confident in the human's intent, cedes control to the human when unconfident, actively requests demonstrations to learn new reward functions to represent novel intents, and adds the newly-learned reward functions to the robot's repertoire for future assistance [9].

While active learning can significantly improve a robot's performance, active queries impose a burden on a human supervisor, as each query interrupts the human and causes them to context switch to the robot's task. To address this concern, I proposed a novel interactive imitation learning algorithm that reduces context switches by a human supervisor by adding hysteresis to the controller that requests human interventions and only asking for help in states that are either novel or high-risk [10, 11]. My research enables a single human supervisor to **simultaneously manage an entire fleet of robots with minimal cognitive workload** [11] and significantly reduces human context switches while achieving higher success rates than prior active imitation learning approaches on complex fabric manipulation tasks [10].

## 2 Research Agenda

My prior work has focused on probabilistic performance bounds in the context of reward functions learned from demonstrations and preference rankings. My research agenda is to expand my prior research along three main axes.

**Leveraging Multi-Modal Human Input** Humans naturally use a variety of modalities when learning from and teaching each other. My prior work has investigated learning from demonstrations [6, 12, 9, 10] critiques [12], and preference rankings [1, 2, 5]. In the future, I plan to develop algorithms that can simultaneously fuse information from a wider variety of human inputs including emergency stops, corrections, natural language, gestures, and gaze. I will develop algorithms that use multiple forms of human input to better reduce uncertainty and more efficiently optimize robust policies. I also plan to develop a mathematical theory of the cognitive complexity and reliability of different forms of human input and use this theory to develop active learning algorithms that allow a robot to **actively choose which type of human input to query** (e.g. demonstration, comparison, e-stop, etc) in order to maximize the amount of information the robot can gain about the human's reward function, while also calibrating the human's proficiency at providing these different input types.

**Interpretable and Verifiable Robustness** My prior work focuses on robustness in terms of formulating and optimizing probabilistic safety bounds; however, these bounds can be hard to interpret, and robustness is ultimately a human-defined and human-judged quality. In the future, I intend to study how to develop methods for explainable and interpretable AI that can assist humans in verifying the robustness of AI systems. As a first step in this direction, my prior theoretical work on machine teaching studies how a AI system can provide maximally informative demonstrations to efficiently teach a human about its overall behavior [15]. This work provides the first theoretical lower bound on the sample complexity of active reward learning. I also recently proved sufficient conditions for the construction of sample efficient **"drivers tests" for AI systems**: tests that efficiently verify that the policy and learned reward function of an AI system are aligned with a human's values [16]. In the future, I plan to expand and build on my theoretical work to develop practical algorithms that allow robots and other AI systems to efficiently and naturally teach people what they know and what they do not know, use active interventions to discover causal factors behind human input, and enable humans to efficiently verify whether black-box AI systems are aligned with their intent.

**Generalizing to Multiple Forms of Uncertainty** For AI systems to be robust, they must be able to efficiently model and reduce a wide range of different forms of uncertainty, not only uncertainty over their objective function, but also uncertainty over system dynamics, sensor observations, object affordances, optimal actions, human rationality and biases, safety constraints, and even uncertainty over the human's own mental model of the robot. As a first step, I recently extended my prior work on high-confidence performance bounds to develop a state-of-the-art grasping algorithm that uses self-supervised learning to discover how to robustly grasp unknown, possibly adversarial, objects [17]. By efficiently maintaining uncertainty over the object geometry and the robustness of different grasps, my work enables a general-purpose grasping algorithm to quickly adapt to novel out-of-distribution objects while maintaining high-confidence bounds on its performance [18]. I have also recently begun working on the problem of using human input to infer safety constraints on a robot's behavior [19]. In the future, I plan to develop **robot algorithms that use limited human supervision to enable efficient and safe exploration**. I also plan to connect my prior work on learning from novices [1, 2] and skilled teachers [15] to develop AI systems that can efficiently maintain uncertainty over the skill-level and teaching ability of a human supervisor in order to provide better assistance and more accurately infer human intent.

I plan to ground my research over the next **10-15 years** in a range of challenging safety-critical problems, including domestic service robots, warehouse robotics, autonomous driving, and healthcare. I have first-hand experience with the challenges of real-world domestic service robotics through RoboCup@Home, where my team took **third place** it the 2017 Domestic Standard Platform League [20]. In the autonomous driving domain, I have recently explored challenges such as learning cost functions that alleviate intrinsic suboptimalities in model predictive control [21] and enabling efficient and accurate modeling of human drivers using multi-fidelity models of human behavior [22]. In domains such as security, autonomous driving, warehouse robotics, and healthcare, it is common to have multiple robots interacting with multiple-humans. As a step towards these types of multi-agent settings, my prior work has investigated control [23, 24, 25] and observability [26, 27] problems in human interactions with swarms of simple robots. In the future, I am interested in enabling "call-center"-like capabilities for human-robot interaction, where a small team of humans remotely supervises and assists a large fleet of robots performing a wide range of complex tasks in people's homes or on factory floors. In the healthcare domain, I am interested in exploring settings, such as assisted living facilities and hospitals, where a small number of robots need to quickly learn from and assist a wide variety humans. I also plan to extend my recent work on offline imitation learning [4] and robust policy optimization [3] to enable robust offline optimization of dynamic medical treatment regimes.

### Funding

I intend to fund my research through the NSF (CAREER, Robust Intelligence, and the National Robotics Initiative 2.0); faculty research awards from industry (Google, Toyota, and Microsoft); and defense funding opportunities through AFOSR (YIP and Computational Cognition and Machine Intelligence) and ONR. **My prior involvement in both writing and reviewing grant proposals while working for the Air Force Research Lab, including writing a successful AFOSR grant, prepare me well to secure similar funding in the future.**

## 3 Conclusion

I envision a world where a robot can observe demonstrations of how I like my dishwasher loaded and know when it has received enough demonstrations to safely perform the task, where AI healthcare assistants combine small amounts of expert queries with large offline datasets to optimize treatment plans that are robust to spurious correlations and ambiguities, and where self-driving cars automatically identify unsafe edge-cases and actively request new data to improve their performance. Successfully and efficiently integrating human input into the study of robust machine learning and robotics will require developing new theoretical and algorithmic techniques and will benefit from insights from fields such as human-factors, causal inference, cognitive science, robust control, and formal verification. I am excited by the many opportunities for collaboration as I work towards the goal of building robust AI systems that achieve, safe, reliable, and beneficial outcomes for society.

# References

[1] **Daniel S. Brown***, Wonjoon Goo*, Prabhat Nagarajan, and Scott Niekum. Extrapolating beyond suboptimal demonstrations via inverse reinforcement learning from observations. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019.

[2] **Daniel S. Brown**, Wonjoon Goo, and Scott Niekum. Better-than-demonstrator imitation learning via automatically-ranked demonstrations. In *Proceedings of the 3rd Conference on Robot Learning (CoRL)*, 2019.

[3] Zaynah Javed*, **Daniel S. Brown***, Satvik Sharma, Jerry Zhu, Ashwin Balakrishna, Marek Petrik, Anca D. Dragan, and Ken Goldberg. Policy gradient bayesian robust optimization. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

[4] Daniel Shin, **Daniel S. Brown**, and Anca D. Dragan. Offline preference-based apprenticeship learning. *Workshop on Human-AI Collaboration in Sequential Decision-Making (ICML)*, 2021.

[5] **Daniel S. Brown**, Russell Coleman, Ravi Srinivasan, and Scott Niekum. Safe imitation learning via fast bayesian reward inference from preferences. In *International Conference on Machine Learning (ICML)*, 2020.

[6] **Daniel S. Brown** and Scott Niekum. Efficient probabilistic performance bounds for inverse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[7] **Daniel S. Brown** and Scott Niekum. Toward probabilistic safety bounds for robot learning from demonstration. In *AAAI Fall Symposium on AI for HRI*, 2017.

[8] **Daniel S. Brown**, Scott Niekum, and Marek Petrik. Bayesian robust optimization for imitation learning. In *Neural Information Processing Systems (NeurIPS)*, 2020.

[9] Matthew Zurek, Andreea Bobu, **Daniel S. Brown**, and Anca D Dragan. Situational confidence assistance for lifelong shared autonomy. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[10] Ryan Hoque, Ashwin Balakrishna, Carl Putterman, Michael Luo, **Daniel S. Brown**, Daniel Seita, Brijen Thananjeyan, Ellen Novoseller, and Ken Goldberg. Lazydagger: Reducing context switching in interactive imitation learning. In *IEEE Conference on Automation Science and Engineering (CASE)*, 2021.

[11] Ryan Hoque, Ashwin Balakrishna, Ellen Novoseller, , Albert Wilcox, **Daniel S. Brown**, and Ken Goldberg. Thriftydagger: Budget-aware novelty and risk gating for interactive imitation learning. In *5th Annual Conference on Robot Learning (CoRL)*, 2021.

[12] **Daniel S. Brown***, Yuchen Cui*, and Scott Niekum. Risk-aware active inverse reinforcement learning. In *Proceedings of the 2nd Annual Conference on Robot Learning (CoRL)*, 2018.

[13] **Daniel S. Brown**, Jeffrey Hudack, Nathaniel Gemelli, and Bikramjit Banerjee. Exact and heuristic algorithms for risk-aware stochastic physical search. *Computational Intelligence*, 2016.

[14] **Daniel S. Brown**, Steven Loscalzo, and Nathaniel Gemelli. k-agent sufficiency for multiagent stochastic physical search problems. In *International Conference on Algorithmic DecisionTheory*, pages 171–186. Springer, 2015.

[15] **Daniel S. Brown** and Scott Niekum. Machine teaching for inverse reinforcement learning: Algorithms and applications. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

[16] **Daniel S. Brown***, Jordan Schneider*, Anca D. Dragan, and Scott Niekum. Value alignment verification. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

[17] Michael Danielczuk, Ashwin Balakrishna, **Daniel S. Brown**, Shivin Devgon, and Ken Goldberg. Exploratory grasping: Asymptotically optimal algorithms for grasping challenging polyhedral objects. In *Conference on Robot Learning (CoRL)*, 2020.

[18] Letian Fu, Michael Danielczuk, Ashwin Balakrishna, **Daniel S. Brown**, Jeffrey Ichnowski, Eugen Solowjow, and Ken Goldberg. Legs: Learning efficient grasp sets for exploratory grasping. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2022. (in submission).

[19] Dimitris Papadimitriou, **Daniel S. Brown**, and Usman Anwar. Bayesian inverse constrained reinforcement learning. In *Workshop on Safe and Robust Control of Uncertain Systems (NeurIPS)*, 2021.

[20] Yuqian Jiang, Nick Walker, Minkyu Kim, Nicolas Brissonneau, **Daniel S. Brown**, Justin W. Hart, Scott Niekum, Luis Sentis, and Peter Stone. Laair: A layered architecture for autonomous interactive robots. In *AAAI Fall Symposium on Reasoning and Learning in Real-World Systems for Long-Term Autonomy*, October 2018.

[21] Avik Jain, Lawrence Chan, **Daniel S. Brown**, and Anca D Dragan. Optimal cost design for model predictive control. In *Learning for Dynamics and Control (L4DC)*, 2021.

[22] Arjun Sripathy, Andreea Bobu, **Daniel S. Brown**, and Anca D Dragan. Dynamically switching human prediction models for efficient planning. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[23] **Daniel S. Brown**, Sean C Kerman, and Michael A Goodrich. Human-swarm interactions based on managing attractors. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2014. **(best paper finalist)**.

[24] **Daniel S. Brown**, Michael A Goodrich, Shin-Young Jung, and Sean Kerman. Two invariants of human-swarm interaction. *Journal of Human-Robot Interaction*, 5(1):1–31, 2016.

[25] **Daniel S. Brown**, Ryan Turner, Oliver Hennigh, and Steven Loscalzo. Discovery and exploration of novel swarm behaviors given limited robot capabilities. In *Distributed Autonomous Robotic Systems (DARS)*. 2018. **(best paper finalist)**.

[26] **Daniel S. Brown** and Michael A Goodrich. Limited bandwidth recognition of collective behaviors in bio-inspired swarms. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems (AAMAS)*, 2014.

[27] Matthew Berger, Lee M Seversky, and **Daniel S. Brown**. Classifying swarm behavior via compressive subspace learning. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5328–5335. IEEE, 2016.