Interdisciplinary Research (IDR) Origination Awards
Cover Page

Project Title

AI-Assisted Decision-making for Stochastic Spatial Tasks with Imperfect Execution

Principal Investigator(s) (full-time faculty)

Name	Department	College
Chris Archibald (PI)	Computer Science	Computational, Mathematical, and Physical Sciences
Nathan Sandholtz	Statistics	Computational, Mathematical, and Physical Sciences
Stefania Ashby	Psychology	Family, Home, and Social Sciences

Track

Track two

Abstract

To date, the field of AI-assisted decision-making (AIDM) has almost exclusively been studied in the context of fixed-response, discrete-action-space tasks. Many important real-world settings do not fit this mold, including decisions which are continuous in nature, and for which the outcomes are stochastic. We refer to these types of tasks as *stochastic spatial tasks*, or SSTs. Decisions in SSTs are not made from a small number of discrete choices, but instead from a continuous action space. We are interested in SSTs undertaken by humans whose actions are corrupted by some degree of *execution error* (i.e., an imperfect ability to exactly carry out their intended actions). Our long-term goal is to create well-founded, evidence-based methods for effective, personalized AI-assisted decision-making in SSTs.

The first steps toward achieving this goal are the development of accurate models of human behavior in SSTs without AI assistance and an exploration of how humans respond to different modes of AI assistance in SSTs. A firm understanding of both of these aspects of human behavior in SSTs is prerequisite to developing AI to aid their decision making. The objective of this IDR proposal is to secure funding for three distinct pilot studies aimed at isolating different aspects of human decision-making and responses to AI in SSTs. Our research team collectively has expertise in psychology, statistical modeling, and artificial intelligence, which uniquely qualifies us to accomplish the proposed work.

Summary of Plans for External Funding

We have identified three NSF programs: *Science of Learning and Augmented Intelligence*, *Research on Innovative Technologies for Enhanced Learning*, and *Human-Centered Computing*, which have goals that overlap with the aims of our project. In February 2024 we submitted a proposal to the first of these programs, *Science of Learning and Augmented Intelligence*, as it seemed to be the most aligned with our research ideas.

While that proposal was unsuccessful, the reviews were encouraging and provided actionable feedback. The predominant theme in the feedback was the need for a tighter scope and the inclusion of detailed experiments. We plan to improve the proposal by incorporating these suggestions and including preliminary results generated from this IDR grant. We will resubmit to this program in August 2025 and continue submitting annually to an NSF program until successful.

Project Narrative

Consider the challenges people face as imperfect executors of decisions in spatial domains:

- A farmer must choose how to fertilize her fields, but may not do so evenly or to ideal concentration. Effectiveness can be measured at harvest, though uncontrollable factors like rain may distort results.
- Military personnel are tasked with selecting airstrike targets, aware that inflicted damage can vary significantly. Myriad sources of uncertainty influence strike outcomes, such as unknown target resilience and potential collateral civilian damage, complicating the decision-making process.
- A tennis player must decide where to aim when serving the ball. His strategy evolves based on his understanding of his own accuracy as well as the point outcomes from previous serves. However, each point outcome depends on the variable actions and mistakes of both players.

Each of these examples involve humans making spatially continuous decisions, where the outcomes of those decisions are inherently stochastic. This randomness can be due to their own inability to execute the chosen action perfectly and/or the influence of other factors beyond their control. We are motivated by the following question: **how can artificial intelligence (AI) help humans make better decisions in these stochastic spatial domains?** While significant progress has been made in understanding how AI can be used to assist human decision-making in some settings, important gaps remain. Human/AI trust dynamics have been studied extensively, but almost always in the context of basic classification tasks [4, 5, 7, 8, 11, 14]. In these tasks a decision-maker chooses between a discrete number of choices (usually two) and feedback is generally "supervised", meaning a binary indication about the correctness of the decision is given. Many potential applications of AI-assisted human decision-making do not fit this classification framework. In particular, there is a class of decisions that, to our knowledge, has not yet been addressed in the research on AI-assisted decision-making: *spatial decisions with imperfect human execution and stochastic evaluative feedback*.

The spatial decisions we are concerned with are not made from a set of discrete choices, but instead are selected from a continuous space. These decisions correspond to target actions that, due to human error, cannot be executed perfectly. Because the set of actions is continuous, the *execution error* induced by the human's imperfect ability is often a complex quantity to model and can be unique for each individual. The imperfectly-executed action yields a numeric evaluation, or reward, from the environment. In contrast to supervised feedback, this *evaluative feedback* requires context in order to ascertain true value. Finally, the stochastic element of these decisions denotes that the feedback may vary, even when executed actions are identical. For brevity, we will refer to decisions with *any* of these characteristics as *stochastic spatial tasks*, or **SSTs**. In addition to the examples above, decisions with these characteristics are present in many other domains, including finance, healthcare, and agriculture.

The use of AI throughout society has increased rapidly over the past decade and growth is expected to continue [9]. As usage expands into new domains, understanding the corresponding influence—either for benefit or detriment—that AI can have on human behavior and overall well-being becomes paramount. Such an understanding is currently lacking for SSTs. While AI techniques have been developed by PI Archibald and others that model human error and approximate optimal choices for humans in SSTs [1, 2, 3, 10, 13], any exploration of the impact of AI interventions on human decisions in SSTs is non-existent. The absence of this knowledge could impede the development of effective AI systems tailored to address the complexities of spatial decision-making in real-world contexts. Thus, comprehensive research and analysis are imperative to ensure that AI advancements contribute positively to societal progress while minimizing potential drawbacks.

Our long-term goal is to produce well-founded, evidence-based methods for effective, personalized AI-assisted decision-making in SSTs. A successful AI advisor must understand the capabilities of an individual human and improve the performance of that human through its interactions.

Planned External Funding Proposal The objective of our next planned external funding proposal is to lay the foundation for AI-assisted decision-making in SSTs, taking the first steps towards our long-term goal. Based on feedback from our initial NSF proposal, our next proposal will focus on the following two specific aims in order to lay this foundation, leveraging our capabilities in cognitive neuroscience and human-subject experiment design.

1. **Understanding human decision-making in SSTs** – It is essential that we understand how humans behave in SSTs, and why they behave the way they do, before we try to augment their choices via AI. We plan to use behavioral and fMRI experiments as well as tools from psychology, statistical modeling, and inverse optimization to observe and model human behavior in SSTs.

2. **Understand how humans respond to various AI interventions** – Effective AI must correctly model and predict the human response to potential interventions. We plan to gather data on these responses using behavioral and fMRI experiments. Tools from psychology, statistical modeling, and machine learning will then be used to develop models of human response that can be used to guide future assistive AI development.

Feedback from our initial NSF proposal was positive, particularly regarding our overarching goals, but reviewers emphasized the need for greater specificity in our experimental design and the application of resulting data. Including pilot studies to demonstrate feasibility and research potential will significantly strengthen future proposals. This IDR proposal seeks to generate preliminary data from three pilot studies, which will enhance our external funding prospects and advance our broader research agenda.

Scope and Purpose of this IDR Proposal The aim of this IDR proposal is to design, conduct, and analyze the results of three pilot human-subject experiments in three distinct SST domains:

1. **Spatial Reasoning and Execution Error – Darts**: A familiar and simple SST that supports controlled behavioral experiments in a laboratory setting.

2. **Neural Mechanisms Active in SSTs – fMRI**: A custom, computer-based artificial SST will be integrated with an fMRI machine to measure neural responses as subjects make decisions and respond to AI input.

3. **Expert behavior in SSTs – Collegiate tennis** : A real-world, complex SST that allows us to investigate expert decision-making and AI interaction in high-level human performance.

These three pilot experiments will enable us to gather a variety of data about human decision-making in SSTs, from low-level neural responses to high-level human performance. In addition, the ability to gather observations of average humans as well as domain experts will allow us validate the applicability of any lessons learned to a wide range of human users. Successful completion and analysis of these IDR-supported experiments will strengthen future external proposals in three key ways: 1) Show that we have the ability to conduct experiments in these domains, 2) Demonstrate the potential benefits of AI assistance in SSTs, and 3) Provide examples of the types of models we can create and the analysis that can be performed on the resulting data. The remainder of this document will detail our specific plans for each of the three experiments that will be supported by this IDR proposal.

Experiment 1: Spatial Reasoning and Execution Error in Darts Darts has been used as a domain for many papers that focus on AI and statistical methods for estimating the execution error of humans and determining where they should aim [6, 10, 12, 13]. As a domain for human subject experiments, darts has the advantage of being fairly well-known and simple. Additionally, it does not require a large amount of space and experiments can be conducted indoors. The overall goal of our behavioral experiments in darts is to gather data and then model the decision-making and beliefs of humans in darts. Specifically, we will explore ideas such as: how well do humans understand their own accuracy and skill at darts? How does previous success/failure impact their future decisions and beliefs? Do humans aim at optimal locations, given their ability? How well can they make decisions in new or unknown settings? Answers to these questions will establish the need for AI-assisted decision-making in domains like darts and also provide us preliminary data to begin developing machine-learning models of human decision-making and beliefs in SSTs.

The behavioral experiment we propose as part of this IDR proposal in the domain of darts will focus on understanding people's beliefs in their own ability and accuracy and will proceed as follows: After providing informed consent and some demographic information, most importantly their past experience with darts, subjects will first have an opportunity to throw a set number of warm-up throws at the dartboard. They will be asked to provide the researcher their aiming point for each throw so that their accuracy can be more easily determined. Next, the subject will be guided through a specific accuracy estimation procedure, where an aiming point on the dart board will be provided by the system and the result of their throws recorded. At this point, the AI system will compute an estimate of the subject's accuracy. Each subject will then be shown a series of 4 circles, projected onto and centered on the middle of the dartboard. After being shown each circle, the subject will be asked how many times they think they can land a dart inside the circle out of 10 throws. The subjects will then throw 10 darts, attempting to make it in that circle each time. After each set of 10 throws, the subject will be asked how well they think they did and their satisfaction with their performance. They then will be shown the next circle. This will be repeated for 4 circle sizes. The circle sizes will be chosen so that each subject will have the same probability of hitting inside the circle, based on the personalized accuracy estimate from the AI system. The subjects will first be shown a circle in which they are expected to land 5/10 darts, then one for 8/10, then one for 2/10, and then one for 5/10 again. The order of the 8/10 and 2/10 circles will be randomized across the participants, to see if there is any ordering effect on their later predictions, performance, or other observations. All of the data will be gathered and analyzed to detect any patterns in the guesses subjects made, their actual performance, and their satisfaction levels. Machine learning models will be trained on this data to see if we can predict the data gathered (prediction/performance/satisfaction) given what happened previously.

This experiment is designed to gain insight into peoples' beliefs about their own accuracy in a natural way, and also to gain insight into how those beliefs might change over time. The experiment will show external grant proposal reviewers an example of what further proposed experiments might look like and the methods we will use to extract scientific insights from them, with the end goal of improving human performance in SSTs with AI.

Experiment 2: Neural Mechanisms of Decision-Making in Computerized SSTs In our second proposed experiment, we aim to learn more about the neural mechanisms associated with 1) decision-making in a continuous-action space and 2) decision-making that involves receiving AI suggestions and input. To mimic a complex, continuous-action space in a more controlled laboratory environment, we plan to design, implement, and utilize a video gaming task that can be performed during functional neuroimaging (fMRI).

The video game task will be designed by the research team in collaboration with computer science and engineering students at BYU who can assist with game creation and coding. Participants will play the video game task on an MRI-compatible video game controller while undergoing fMRI at the BYU MRI Research Facility. The planned task will provide participants with subtasks or goals to accomplish throughout gameplay with a continuous range of decisions that can be made in order to accomplish a single goal. For some sub-goals participants will be left to their own decision-making processes, so we can measure neural activity during decision-making in a continuous-action space. For other sub-goals, participants will be provided with AI-generated suggestions for accomplishing the task more efficiently and accurately.

We are primarily interested in examining neural activity under three decision-making conditions: 1) when participants are generating their own decisions, 2) when participants are following AI input or suggestions, and 3) when participants choose to ignore AI input and suggestions and go their own direction. Additionally, we will examine behavioral metrics of performance such as reaction times, time to complete task sub-goals, and general accuracy across the same three decision-making conditions. Importantly, the gaming environment will allow us to look at small and complex sequences of decisions in a continuous-action space during fMRI and to learn more about the neural mechanisms that support decision-making in an SST with and without AI input.

The funds requested as part of this IDR proposal will support development of the video gaming task as well as collection of data with a handful of subjects to use as pilot data that will allow us to demonstrate our fMRI capabilities and motivate specific task refinements and future work as we seek additional funding through the NSF.

Experiment 3: Serving Decisions in Collegiate Tennis In our third proposed experiment, we aim to collect data in an SST in which the decision-makers are trained experts in the task. In collaboration with the BYU men's and women's tennis teams, we will conduct an experiment focused on serving decisions in tennis. Much of the groundwork for this experiment is laid; we have connections with the tennis coaches and we have already gathered an initial data from eight current and former players from the men's team.

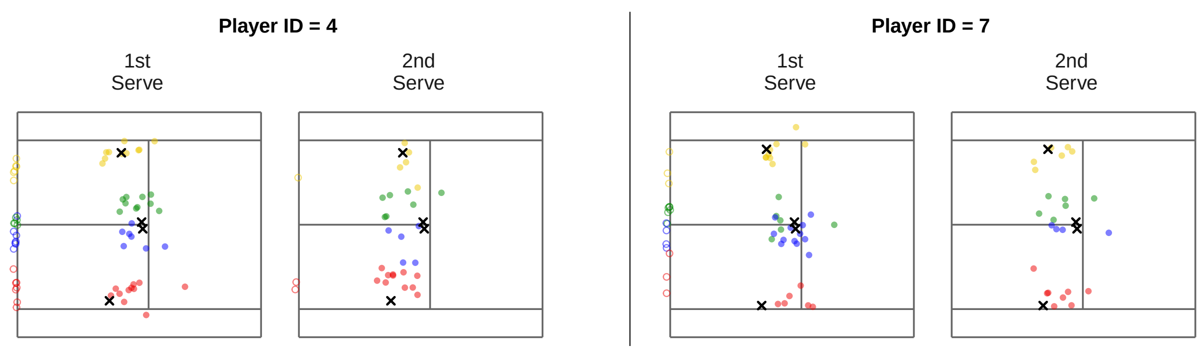


Figure 1: *Observed data for two players from our initial tennis experiment. The black x's show the players' chosen target locations. Filled colored dots show observed bounce locations and unfilled dots represent serves that hit the net.*

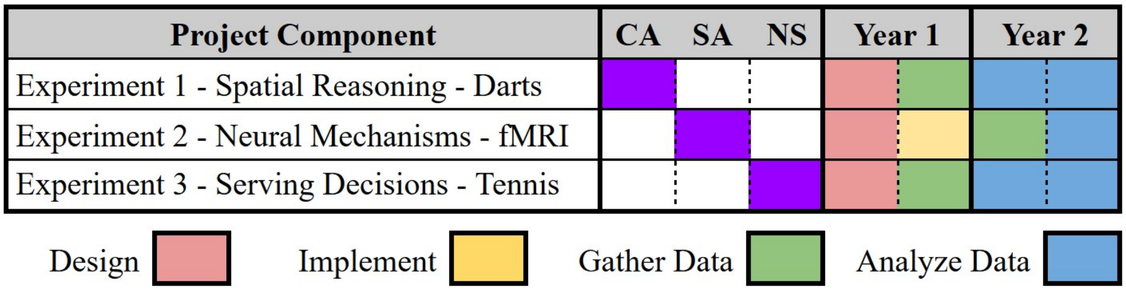
For each participating athlete, the experimental setup consists of two main phases: target selection and data collection. In the target selection phase, the player identifies the precise aiming locations they believe to be their optimal serving targets. For both the ad and deuce courts, they designate an "out-wide" target, positioned near the singles boundary, and an "up-the-T" target, located adjacent to the mid-court line (see the black "x"s in Figure 1). These chosen aiming locations are then marked with tape on the court, ensuring that the player can see their targets when serving.

During data collection, the player delivers 60 serves to a returner, with an equal split between the deuce and ad courts. For each serve, the target—either "wide" or "T"—is randomly assigned and communicated to the server. The returner remains unaware of this information. The player serves and the point is played out. Serve speed is measured using a radar gun. Before each serve, the ball is lightly misted with water which creates a visible mark where it bounces on the returner's court, allowing for precise spatial tracking. After the point, a recorder places a numbered sticker on the water mark corresponding to the serve attempt index. Once the session is complete, the recorded bounce locations are documented in polar coordinates.

Upon collecting this data, we will develop player-specific execution error models, which then will be utilized to estimate optimal strategies via AI. In a follow-up experiment, we will communicate these optimal strategies back to the players and track how their behavior changes.

These experiments will provide insights into how experts make strategic decisions in SSTs and the most effective ways of aiding their decision making via AI. Furthermore, the findings from this research could inform the development of AI decision-support in similar domains.

Investigator Roles and Timetable This table indicates the specific aims where each investigator (indicated by their initials) will have a primary role in accomplishing the proposed work. In addition, the table shows what will be done at each stage of the project.



PI Chris Archibald has expertise in AI techniques for execution error modeling and decision-making in SSTs, and will lead experiments and analysis in the domain of darts. **Co-PI Stefania Ashby** is experienced with exploring factors that influence human memory and decision-making and designing behavioral and functional neuroimaging experiments that give insight into those factors. She will be involved with the cognitive psychology aspect of the project overseeing collection of cognitive measures as well as neuroimaging experimental design and data collection. **Co-PI Nathan Sandholtz** has experience using inverse optimization techniques to model human decisions and will lead efforts in the domain of tennis, applying those techniques to the data obtained.

Expected Outcomes **1) External funding proposals** - the work accomplished as part of this project will provide crucial preliminary results to bolster and improve future proposals for external funding from the NSF. **2) Scientific knowledge and research publications** - the proposed work will add to our understanding of many topics related to AI-assisted decision-making in SSTs. This new knowledge will be shared by publication in conferences and journals within the statistics, AI, and psychology academic communities.
3) Student mentoring - the undergraduate students that will be supported by this grant will be mentored by the faculty investigators.