Proceedings of the Nineteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2023)

Risk Management: Anticipating and Reacting in StarCraft

Adam Amos-Binks[1*], **Bryan S. Weber**[2]

[1] Applied Research Associates Inc., Raleigh, NC 27616 USA
[2] College of Staten Island: CUNY, SI, NY 10314 USA
aamosbinks@ara.com, bryan.weber@csi.cuny.edu

Abstract

Managing risk with imperfect information is something humans do every day, but we have little insight into the abilities of AI agents to do so. We define two risk management strategies and perform an ability-based evaluation using StarCraft agents. Our evaluation shows that nearly all agents mitigate risks after observing them (react), and many prepare for such risks before their appearance (anticipate). We apply traditional causal effect inference and causal random forest methods to explain agent behavior. The results highlight different risk management strategies among agents, strategies that are common to agents, and overall encourage evaluating agent risk management abilities in other AI domains.

Introduction

Managing risk with imperfect information is required in daily life (e.g., bringing an umbrella when it might rain). AI agents do the same, from autonomous vehicle controls that slow down when vision conditions deteriorate to games such as StarCraft where a partially observable environment ensures agents have imperfect information about the game state. Agents are adept at managing risks under these conditions, evidenced performance that approaches human-level in everyday tasks such as driving and game-playing.

Despite the prevalence and necessity of managing risk in these AI tasks, risk management is not evaluated directly, and we know little of its relationship to task performance (Amos-Binks, Dannenhauer, and Gilpin 2023). Existing AI domains have been almost exclusively focused on task-based evaluations. The ability to beat and compete with world-class players (e.g., Go (Silver et al. 2017), StarCraft (Vinyals et al. 2019)) and question-answering (e.g., Winograd schemas (Levesque, Davis, and Morgenstern 2012)) only use the final outcome in their evaluations. Ability-based evaluations that cut across domains have received comparatively little attention. Universal psychometrics (Hernández-Orallo et al. 2017) is an active area of research but has yet to find a foothold among existing AI tasks. Finally, the explainable AI community has endeavored to provide machine learning model user's with insights into *why* a model made a specific prediction (Gunning et al. 2021). As AI continues to pervade everyday life, such as in mission-critical systems, we require even more transparency into an agent's risk management abilities.

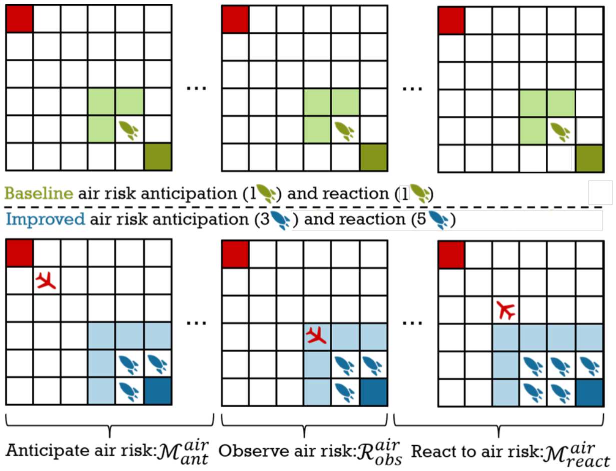
Baseline air risk anticipation (1) and reaction (1)
Improved air risk anticipation (3) and reaction (5)
Anticipate air risk: $\mathcal{M}^{air}_{ant}$
Observe air risk: $\mathcal{R}^{air}_{obs}$
React to air risk: $\mathcal{M}^{air}_{react}$

Figure 1: We calculate the blue agent's ability to manage risks with imperfect information both before observing a risk (anticipation) and after (reaction). We adjust these calculations using a baseline so we can determine whether agents under- or over-perform.

Our approach tackles the lack of risk management characterization with an ability-based evaluation of risk management in StarCraft AI agents. StarCraft provides an ideal domain for evaluating risk management as defender agents are unaware of when and to what degree (imperfect information due to fog-of-war) attacker agents develop capabilities (risks such as air, cloaked units) for which "hard counters" are developed (mitigations such as anti-air, detector units) to perform well on the overall task (win).[1] Figure 1 depicts a simplified example of air units in red and anti-air units in green/blue. Our study makes three contributions. First, we extract and share the relevant data for risk management from the AIIDE 2020 StarCraft tournament, the most recent at the time of writing. The diversity of agents and tournament design (many games, time limits, known maps, fog-of-war) are

[*] Contact author

[1] A "hard counter" is a unit that explicitly defeats or undermines another unit.

ideal for evaluating risk management. Second, we define the risk management terms needed to perform our evaluation. While we evaluate StarCraft agents, we expect the definitions for calculating anticipation and reaction strategies can be adapted for other tasks to evaluate risk management ability. Finally, we evaluate the abilities of tournament agents to manage risk. To do this, we use a novel application of traditional causal inference methodologies that compares a baseline counterfactual against an agents actions, Figure 1. We explain an agents' behavior as "reacting" or "anticipating" to risks, an explanation that requires identifying the furtive cause of the behavior. We find that many agents react to observing risks by investing in mitigations. We also found that the average tournament-level agent anticipates risk by investing in mitigations before risks are even observed.

Background

Our contributions build on three different background areas. First, StarCraft is a real-time strategy game (RTS) with properties conducive to applying our risk management evaluation. Second, our evaluation is best characterized as an ability-based AI evaluation. We deliberately characterize it as an ability-based evaluation because we intend risk management abilities can be assesses in other domains, unlike task-based evaluation that are domain specific. Finally, we discuss the concepts we employ from risk management.

StarCraft RTS

While the original StarCraft was released in 1998, it has become a focal point of AI research with many tools, tournaments, and agents in active development. StarCraft's fog-of-war mechanic, where players only have visibility to game tiles immediately near their units (imperfect information), along with the known capabilities opponents can develop (risks), and countering capabilities (risk mitigations), are ideal properties for assessing risk management. We use these properties to measure the change in agent resource allocation before and after observing an opponent's threatening capability.[2] We focus on the AIIDE StarCraft AI competition, run annually with the AIIDE conference, as tournament matches contain fog-of-war, there is a range of AI agents, and the tournament design provides a high quality data source as agents play over 100 matches per opponent.

AI Evaluation

AI evaluations are broken down into task- and ability-based evaluations (Hernández-Orallo 2017). Task-based are solely concerned with final task performance without investigating the abilities to achieve them. Our survey of StarCraft evaluations found only task-based evaluations using StarCraft match winners (e.g. (Churchill, Buro, and Kelly 2019)) or overviews of the tournaments in general (Čertický and Churchill 2017; Čertický et al. 2018). In contrast, ability-based evaluations focus on characteristics that enable performance across several tasks. Progress in exploring universal psychometrics is promising but not concrete enough to apply directly to StarCraft. Instead, we focus on evaluating risk management from the ground up in StarCraft, aiming to demonstrate the effectiveness of modern causal inference techniques and inspire their application to other tasks.

Risk Management

Typical risk management involves five steps; identify, analyze, prioritize, mitigate, and monitor. In many domains (e.g., insurance), risks are already identified, we view StarCraft as one of these domains. Analyzing and prioritizing within StarCraft amounts to enacting mitigations and observing a risk in a match. Mitigation is a measure of the degree of mitigation investment, while monitoring ends when the match does. We formalize risk management approaches with definitions for *anticipation* (risk mitigation investment before observing risk) and *reaction* (risk mitigation investment after observing risk) in Section .

Here, we are attempting to disambiguate if a risk actually causes an agent to respond, or if the various motions of an agent are merely correlated and synchronistic. One such explicit example these tools help disambiguate would be that agent may blindly build anti-air at a particular frame count every game because of build order (in which case the active threat does not cause anything *per se*), from whether or not they are actually responding to a threat observed (and therefore the threat causes a response). We do this by comparing a counterfactual state (games where no such threat appears) to the actual state (games where threats do appear) and contrasting the average of weighted subsets of these games by several well-known techniques in Section . In other fields, researchers are interested in if a particular action causes a response, such as if adding a minimum wage affects unemployment. The parallels are direct - the "threat" is adding the minimum wage and the response (by the economy at large) is the changing unemployment level (Card and Krueger 1994). Causal effect tools are for measuring these cases and have wide reach, for example this particular example has about 5000 citations.

Risk Management Abilities

In this section, we provide definitions that culminate in the two risk management strategies: anticipation and reaction. After each definition, we describe a concrete, illustrative StarCraft example that also segways into our Evaluation in Section . Our definitions are specific to StarCraft, though we expect they are general enough to support risk management ability evaluations in other real time strategy games and domains with mechanics that have 'hard counters'.

In our example, we refer to two agents playing each other as the *defender* in blue and the *attacker* in red. StarCraft capabilities vary by race but are static and well-known as the game is no longer under active development. We define attacker unit types that have a specific capability as a risk:

Definition 1 (Risk) *The set of attacker unit types with a capability cap is $\mathcal{R}^{cap}$ and the set of attacker unit instances produced in a match up to a given time t is $\mathcal{R}_t^{cap}$.*

[2] In an interview, the author of the McRave StarCraft agent described our definition of reaction as what practitioners would call "panic buying".

Our example in Figure 1, column 1, row 2, the attacker's air units present a risk as they have the unique ability to inflict damage by shooting 'down'. For a defender to mitigate the risk of air units, the defender agent needs unit types with the ability to inflict damage by shooting 'up' (ground units) or 'across' (air unit). We define risk mitigations as:

Definition 2 (Risk Mitigation) *Defender units with the capability $\mathcal{M}^{cap}$ necessary to inflict damage to any unit in $\mathcal{R}^{cap}$. All defender unit instances with a type in $\mathcal{M}^{cap}$ up to a time t of the current match is referred to as $\mathcal{M}_t^{cap}$ where m_i contains the number of units produced of that unit type.*

In Figure 1, column 1, row 2, the defender's air mitigations ($\mathcal{M}_{obs}^{air}$) are three anti-aircraft units (blue rockets). These mitigating units reduce the risk posed by an attacker's air units. We calculate a defender's risk mitigation investment using an existing aggregate resources value (Synnaeve and Bessiere 2012; Weber 2018) for $\mathcal{V}^{cap}$:

Definition 3 (Mitigation Investment) *The total quantity of resources invested to mitigating risk cap up to time t is calculated as the dot product of the value vector $\mathcal{V}^{cap}$ that contains the resources required to produce unit types with cap and the Risk Mitigation vector $\mathcal{M}_t^{cap}$ for cap at t:*

$$Inv(\mathcal{M}_t^{cap}, \mathcal{V}^{cap}) = \sum_{i}^{|\mathcal{M}_t^{cap}|} m_i v_i \tag{1}$$

In Figure 1 column 1, row 2, the defender has invested in 3 mitigating units (blue rockets, $\mathcal{M}_{obs}^{air}$), if their cost is \$2 each ($\mathcal{V}^{air}$), then they have invested $Inv(3, 2) = \$6$. While risks (capabilities) are known *a priori* in each match, attackers acquire them through a series of building upgrades and intermediate capabilities. It is not clear the risks an attacker will develop or when they may do so. In StarCraft, the fog-of-war game mechanic grants only imperfect information about when their attacker has produced a given risk (capability). We define when a defender observes the first attacker's unit that poses risk $\mathcal{R}^{cap}$ (risk unit first present in a visible game tile) as an Observed Risk.

Definition 4 (Observed Risk) *The first time an attacker's unit with type in $\mathcal{R}^{cap}$ is located in a game tile visible to a defender in the current match is $\mathcal{R}_{obs}^{cap}$*

In Figure 1, the time a defender observes attacking air units is column 2, row 2 with risk value $\mathcal{R}_{obs}^{air} = 2$, if we assume one air unit has a nominal value of \$2. Denoting when a defender first observes a risk is essential as it allows us to measure how effectively a defender manages imperfect information about risks. Drawing on inspiration from anticipatory thinking (Geden et al. 2019), we refer to a defender's investment in risk mitigations **before** observing a risk as Anticpation:

Definition 5 (Anticipation Investment) *The quantity of resources a defender has invested in risk mitigations before $\mathcal{R}_{obs}^{cap}$ is $Inv(\mathcal{M}_{ant}^{cap}, \mathcal{V}^{cap})$ such that $ant <= obs$.*

We calculate the defender's anticipation investment for air risks in Figure 1 as $Inv(\mathcal{M}_{ant}^{air}, \mathcal{V}^{air}) = 1.5$ where, for illustration, $\mathcal{V}^{air} = [0.5]$ and there are 3 defending rockets. To support evaluating whether an agent is truly anticipating and preparing for a risk, we adjust a defender's anticipation investment by subtracting a counterfactual baseline amount to obtain Anticipation:

Definition 6 (Anticipation) *The difference between anticipation investment (Definition 5) and a counterfactual where the risk is never observed, $\mathcal{R}_{obs}^{cap} = 0$, is the anticipation, $Inv(\mathcal{M}_{ant}^{cap}, \mathcal{V}^{cap})|\mathcal{R}_{obs}^{cap} > 0 - Inv(\mathcal{M}_{ant}^{cap}, \mathcal{V}^{cap})|\mathcal{R}_{obs}^{cap} = 0$*

In our example, the anticipation in Game 2 is 2, the difference between the 3 blue and 1 green anti-air units from the baseline. Observing an attacker's capability prompts a defender to react. The defender must still account for the imperfect information of the strength/degree of the risk (i.e., the number of units an attacker has that pose a risk). Reaction is a defender's investment in risk mitigations after observing a risk:

Definition 7 (Reaction Investment) *The quantity of resources a defender has invested in risk mitigations after $\mathcal{R}_{obs}^{cap}$ is $Inv(\mathcal{M}_{react}^{cap}, \mathcal{V}^{cap})$ such that $obs < react$.*

After observing the attacker's air risk in row 2, col 2 of Figure 1, we calculate the defender's reaction investment as $Inv(\mathcal{M}_{reac}^{air}, \mathcal{V}^{air}) = 1$ where $\mathcal{V}^{air} = [0.5]$. This formula is similar to the definition of Anticipation Investment, but with a change in timing to occur after the threat appears instead of before (i.e. $ant <= obs < react$). To support our evaluation of reacting to observing a risk, we adjust a defender's reaction investment by subtracting a counterfactual baseline amount to obtain Reaction:

Definition 8 (Reaction) *The difference between the reaction investment (Definition 7) and a counterfactual where the risk is never observed, $\mathcal{R}_{obs}^{cap} = 0$, is the reaction, $Inv(\mathcal{M}_{react}^{cap}, \mathcal{V}^{cap})|\mathcal{R}_{obs}^{cap} > 0 - Inv(\mathcal{M}_{react}^{cap}, \mathcal{V}^{cap})|\mathcal{R}_{obs}^{cap} = 0$*

In the example in Figure 1, the reaction in Game 2 is 2, the difference between the 3 blue and 5 blue anti-air units in columns 2 and 3, respectively.

Both reaction and anticipation are calculated on a per match basis. In the next section, we use anticipation and reaction calculations to characterize an agent's ability to manage risk with imperfect information over multiple matches.

Evaluation

StarCraft provides an ideal domain to evaluate ability-based risk management strategies. The risks and mitigations are widely known. This ensures defenders are managing known risks and not employing strategies that might be effective in an open-world where novelty exists.

We focus on air and cloaking risks as they have what many consider "hard-counters", allowing us to evaluate agent risk management strategies. We refer to anti-air units as those that can inflict damage to air units, a capability that not all ground units possess. Anti-cloak units are those that reveal the location of an attacker's cloaked units, which is necessary to inflict damage on them. We note there is contextual variability in how suitable each unit in this list is for the task of countering, but this is the exhaustive list of such units. There is also an argument to be made that ignoring a threat is sometimes the best strategy - but we are specifically exploring and measuring anticipation and reaction, not optimal strategy selection as in (Ballinger and Louis 2013; Buro et al. 2012; Churchill et al. 2011; Ontañón et al. 2013). Table 4 in Appendix , details the aforementioned risks across all StarCraft races. We assume agents will respond to air and cloak risks with by anti-air units or detectors.

We use the 2020 AIIDE StarCraft AI Tournament for the diverse skill level of agents, replays to extract content from, and many matches between all tournament participants to ensure robust statistical results.

Evaluation Data

The AIIDE 2020 StarCraft AI Tournament consists of thirteen agents that play one hundred and fifty games (slightly less in some cases due to technical difficulties) against all other agents. We use the BroodWar API (Heinermann et al. 2012) in a custom program to extract game state data from 11695 game replays into csv format.

This data is summarized in Table 1.

	Total Value of:			Rounds After:	
	All Units	Air Units	Cloak Units	Air Threats	Cloak Threats
Mean	7305	1324	226	27.4%	22.2%
Std	8087	2930	547	44.6%	41.5%

N:240,023

Table 1: Summary of in-game units

Every 30 seconds (720 frames) we take inventory of all units on the board for both participants. We drop all very early data prior to any possible air or cloak attack (before 6000 frames or 4 minutes), for a total of 240,023 examined frames. We simplify this inventory by calculating the net worth of all units of each risk mitigation by summing the point value for each unit (point values are effective for training AI agents (Synnaeve and Bessiere 2012; Weber 2018)).

We found that 27% of frames occur after the first air unit is created, and 22% of frames occur after the first cloaked unit is created, suggesting both of these risks are used regularly.

Study Design

We evaluate the association between a defender's first observation of a risk unit at time t for agent i in game g (indicated by $\Sigma_{j=0}^{t}\mathcal{R}_{ijg}^{cap} > 0$, now referred to as W_{ijg} for brevity), and the log value of total risk mitigation units indicated by $ln(\mathcal{M}_{itg}^{cap})$.[3] We refer to each observation using x_{itg} instead of x_{obs}. Specifically, we are attempting to measure: $E(ln(\mathcal{M}_{itg}^{cap})|W_{ijg} = 1, A) - E(ln(\mathcal{M}_{itg}^{cap})|W_{ijg} = 0, A)$, the difference between the amount of $ln(\mathcal{M}^{cap})$ in cases where the agent has observed the risk and the counterfactual case where the risk was not observed and all else (A) being held equal.

The method of performing this evaluation is called "difference-in-differences" (DiD). Recall the simplified pair of three-period games shown in Figure 1 where row 1 is a baseline game 1 in which air risk ***was not*** observed by the defender, and row 2 was a different game 2 where an air risk ***was*** observed by some other defender. Column 3 indicates frames after air units would have shown up (if they did). Game 1 uses its previous states as a baseline, $E(\mathcal{M}_{t=3,g=1}^{cap}|W_{t=3,g=1} = 0) - E(\mathcal{M}_{t<3,g=1}^{cap}|W_{t<3,g=1} = 0) = 1 - 1 = 0$ anti-air units are added at the time of threat. In Game 2, we also use the previous states as a baseline, $E(\mathcal{M}_{t=3,g=2}^{cap}|W_{t=3,g=2} > 0) - E(\mathcal{M}_{t<3,g=2}^{cap}|W_{t<3,g=2} > 0) = 5 - 3 = 2$ anti-air units. We then compare the difference between the reactions of agents in game 1 and 2 as counterfactuals for one another, the size of the reaction is the $2 - 0 = 2$ units, the reactions by the agent in game 2 is estimated to be of size 2, even though the blue agent in game 2 had more investment prior to the threat. In this illustrative example, we can observe that no other relevant factors have changed and assume both agents follow the same parallel trends (Angrist and Pischke 2008). This technique is not a panacea- discussion of the appropriate comparison groups, and adapting to limitations of this technique is an area of active research (Callaway and Sant'Anna 2021).

Methodology: Causal Methods

In the causal effect literature, the stimulating event is referred to as a "treatment", and we are attempting to measure the "effect of treatment" – specifically the effect of each first observation of an attacker's air and cloak units. Our goal is to estimate the effect of treatment (W_{ijg}) on an outcome ($\mathcal{M}^{cap}$) with a set of features X_{ijg} which control for observable characteristics of the game. The difficulty is that treatment is either observed or not - we cannot rewind and fork time to consider both cases for a particular individual and time period. Therefore, it is critical to these causal methods that we construct a counterfactual treatment (or nontreatment) case using other available data.

The typical method of estimating difference-in-differences is designed for data where individuals are observed for a single timespan (there is typically no g (game) dimension to the data) (Angrist and Pischke 2008). Two-way fixed-effect estimation refer to the presence of a set of indicators for both the set of time periods and the set of individuals. We could apply the typical methodology directly if, hypothetically, we had a data from a large number of tournament participants playing one game each and each individual serves as a counterfactual for others. In the AIIDE tournament, we have repeated games - and repeated games by the same agents serve as close counterfactuals. This estimator is called the "three-way fixed-effect estimator," and the general projection that is universally applicable is to include all interactions of each

[3] We use the adjustment $ln(1+\mathcal{M}_{itg}^{cap})$ since sometimes the total value of mitigating units is 0.

of the 3 dimensions (Balazsi, Matyas, and Wansbeek 2017).

$$\begin{aligned} ln(\mathcal{M}_{itg}^{cap}) = \beta_0 + \beta_1 W_{itg} + \beta_2 X_{itg} + \beta_3 L_{itg} \\ + \lambda_{1it}[Individual]_i * [TimePeriod]_t \\ + \lambda_{2tg}[TimePeriod]_t * [Game]_g \\ + \lambda_{3ig}[Individual]_i * [Game]_g + \epsilon_{itg} \end{aligned} \tag{2}$$

Here, treatment is measured by the appearance of a risk: W takes the value 1 if a risk has been observed at that time in the game (or earlier) and 0 otherwise. The coefficient β_1 measures the estimated effect of treatment, it is the difference between the treated and untreated counterfactuals. The covariates, X_{igt} are the defender's value of all assets (which changes within a game) and race (which may change for random players). L is a binary variable which takes the value 1 if it is the period before a risk is observed and 0 otherwise - we include this factor if and only if we are examining anticipation effects, and the coefficient β_3 estimates that anticipation. We include an indicator for each time period t (encoded one-hot, which we indicate with brackets) and an indicator for each observed individual i and game g (also encoded one-hot). These indicators are interacted with one another to capture all possible interactions of game, individual, and time period. These are called "fixed effects" because they account for the average effect of a fixed period, defender, or game. One element (called the base) must be excluded from each dimension to avoid multicollinearity, for a total of $(I-1)*(T-1)+(I-1)*(G-1)+(G-1)*(T-1)$ fixed effects. We examine this model for the two different risks – where *cap* is air and cloaking – and measure both anticipation and reaction using difference-in-differences.

Reaction: Our estimate of reaction is the significance of β_1, where the presence of the treatment is the appearance of a threatening cloak or flying unit indicated by the indicator W_{itg} shifting from 0 to 1. After the threat appears, W_{itg} remains as 1, so β_1 represents the estimated impact of observing a threatening unit relative to an equivalent counterfactual un-threatened period, $E[ln(\mathcal{M}_{itg}^{cap})|W_{itg} = 1] - E[ln(\mathcal{M}_{itg}^{cap})|W_{itg} = 0]$.

Anticipation: Our estimate of reaction is the significance of β_3, where the appearance of a threatening cloak or flying unit in the *next period* is indicated by the indicator L_{itg} shifting from 0 to 1. If the observation of a risk in the next period is a significant predictor of response in period t, then we conclude the agent used some information in period t to anticipate the appearance of a risk in period $t+1$.

Causal Random Forest Our goal is to identify under what conditions each agent is likely to respond (anticipate, react, or not at all) to risks. Since DiD only finds average causal effects, we supplement the above examination with a more recent model, the causal random forest (CRF) (Wager and Athey 2018; Athey and Wager 2019; Athey, Tibshirani, and Wager 2019; Weber and Cappellari 2022). Most important here is the the ability to use CRF to estimate reactivity along specific dimensions of the data, such as an individual players' reactivity as the game progresses.

This model leans on the intuition of propensity score weighting (Hirano, Imbens, and Ridder 2003). The first idea is that one can estimate the predicted outcome by constructing a tree, and evaluating the average outcome within the outcome leaf F using a set of unconfounded features X_i and the observation of a threatening unit W_i to find an estimated outcome, in this case the amount of M^{cap} with estimated value indicated by $\hat{M}^{cap}$.

$$\hat{\mathcal{M}}_{itg}^{cap} = \frac{1}{|\{itg : X_{itg} \in F(x)\}|} \sum_{itg:X_{itg} \in F(x)} \mathcal{M}_{itg}^{cap} \tag{3}$$

Granting that this estimation process is unconfounded (Rosenbaum and Rubin 1983), we can recognize that those in the same leaf have similar propensities for treatment. As such, we can estimate impact of exposure for one tree by $\hat{\tau}$ by subtracting the estimated mitigation amount with and without the threat for all elements in each leaf:

$$\begin{aligned} \hat{\tau}(x_{itg}) = & \\ & \frac{1}{|\{itg:W_{itg}=1, X_{itg} \in F(x)\}|} \sum_{\substack{itg:W_{itg}=1, \\ X_{itg} \in F(x)}} \mathcal{M}_{itg}^{cap} \\ & - \\ & \frac{1}{|\{itg:W_{itg}=0, X_{itg} \in F(x)\}|} \sum_{\substack{itg:W_{itg}=1, \\ X_{itg} \in F(x)}} \mathcal{M}_{itg}^{cap} \end{aligned} \tag{4}$$

As we expand the number of trees B, the average treatment effect is directly estimated by averaging $\hat{\tau}(x_i) = B^{-1}\Sigma_{b=1}^{B}\hat{\tau}_b(x)$ across all trees. The features (X) we use for CRF inputs are the same as for the previous model: defender's net value of all assets, the time period, the defender's name, the defender's race, and the game ID.

Like with difference-in-differences, using W_{itg} provides an estimate of $\hat{\tau}$ measuring the average effect of reaction. If we wish to instead use L_{itg}, we estimate a $\hat{\tau}$ measuring average effect of anticipation.

Estimating heterogeneous treatment effects requires stronger assumptions: honest trees (a construction choice to forego outcome information while estimating the tree) and that treatment and non-treatment cases are sufficiently dispersed throughout the tree (α-regularity in splits). We note we do not have sufficient dispersion of air/cloak threats to evaluate anticipation using CRF, the period before risk exposure tends to arrive at similar time periods in the game with great regularity. This is because the agents tend to execute build orders precisely, and therefore the estimation pivots on the information in a single 30-second time period before the threat arrives, and this period is usually the same every time. On the other hand, in estimating reactions, there are many periods after the threat arrives to serve as contrasts. We already discard extremely early periods where threats are possible. For details of the calculation, proof of consistency, and standard error, see (Wager and Athey 2018).

Results

Risk Management In Aggregate, DiD One can describe the goal of this DiD approach as trying to measure the

”jump” in the response variable after treatment relative to the untreated groups. We have visualized the jump for all agents within games treated by cloak in Figure 2.

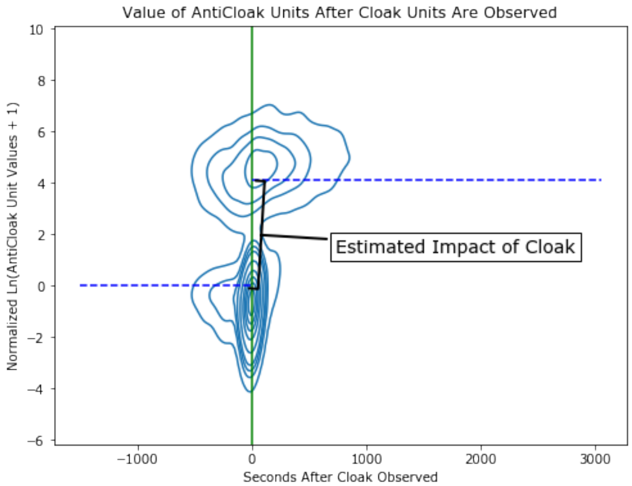

Figure 2: The dotted blue line represents a fit of the log value of anti-cloak units a player has before and after observing air units, normalized around 0. The solid green line, also normalized around 0, represents the moment a cloaking unit is observed (in games where this has occurred). There is a large and significant jump in the log value of anti-cloak units after observing that the enemy team has cloaking units. Normalized for all listed fixed effects and controls.

This figure provides a contour plot of the log value of anti-cloak units before and after a cloaked threat appears. We normalize by the large slate of fixed effects in the regression, so one could equally describe this figure as a residual plot of a regression containing only these factors, following the Frisch-Waugh theorem (Stock and Watson 2012). Figure 2 shows that there is a jump in the value of log anti-cloak units after agents observe cloak units, after controlling for all fixed effects and controls.

The relative increase in mitigating units is measured precisely by estimating Equation 2, and the regression results of that equation are presented in Table 2.

Dependent Variable: $ln(\mathcal{M}_{itg})$

	Anticipate		React	
Coef.	Cloak	Air	Cloak	Air
W_{itg}	4.26***	0.33***	4.13***	0.30***
	(0.01)	(0.02)	(0.01)	(0.25)
L_{itg}	3.07***	0.84***	-	-
	(0.04)	(0.04)	-	-
R^2	0.57	0.33	0.55	0.32
N	240,023	240,023	240,023	240,023

Standard errors, in parentheses, are clustered by player.
Significance levels: ***0.01, **0.05, * 0.1

Table 2: Difference In Differences Regression Results

Table 2, the *React* columns present the estimated DiD coefficient β_1 in row 1, representing the estimated impact of observing a nonzero $\mathcal{R}^{cap}$ on the quantity $ln(\mathcal{M}^{cap}_{itg})$, in other words measuring the average size of “panic buying.” In the case of cloak risks, shown in column 3, we find a significant increase of about 4.13 in the log quantity of resources directed towards anti-cloak units *after* a cloaking unit has been discovered. When adjusted for a log-level perspective, this means that there is approximately a 413% increase in the value of anti-cloak units relative to the baseline after identifying a cloaking risk (Stock and Watson 2012). This means that in frames following the discovery of the cloak units, AIIDE 2020 agents have substantially more units that attack cloak units - relative to the agents' average for that time and type of game.

We repeat the same methodology for the risk of air units and present the results in column 2. The DiD coefficient is a significant 0.30, and under the same interpretation of log-level regression, this means there is a 30% increase in the value of anti-air units after the first observation of an air risk. The significance of the coefficients in Table 2 are consistent with the hypothesis that, on average, the pool of agents significantly responds to both cloak risks and air risks.

In Columns 1 and 2 of Table 2, we explore the idea that there is anticipation by indicating the period before the threat appears, L_{itg}, in row 2. Because this coefficient is significantly positive, we find evidence that agents, on average, appears to have significant buildup of mitigations in the period before observing a risk for both air and cloak, implying anticipation. We also find evidence of significant reaction for both air and cloak units. However, there is variability among the agents, some agents react more than others. We next investigate CRF to determine if it estimates different treatment effects, and to exploit its ability to estimate heterogeneous responses.

Risk Management In Aggregate, CRF Next, we use the causal random forest with the same covariates as the previous estimate.

Dependent Variable: $ln(\mathcal{M}_{itg})$

	Anticipate		React	
Coef.	Cloak	Air	Cloak	Air
$\hat{\tau}$	-603†	-371†	3.58***	3.68**
	(0)	(15014)	(0.97)	(1.35)
N	240,023	240,023	240,023	240,023

Standard errors, in parentheses.
Significance levels: ***0.01, **0.05, * 0.1
†: process indicated challenges, discussion below.

Table 3: CRF Estimation Results

In Table 3 columns 3 and 4, we find a similar story to the DiD results. The average agent has significant reactions to both air and cloaked units (W_{itg}) when controlling for combinations between covariates and the propensity for treatment. Unlike in DiD, the magnitude of the estimated reaction is large for both air and cloak threats. The difference

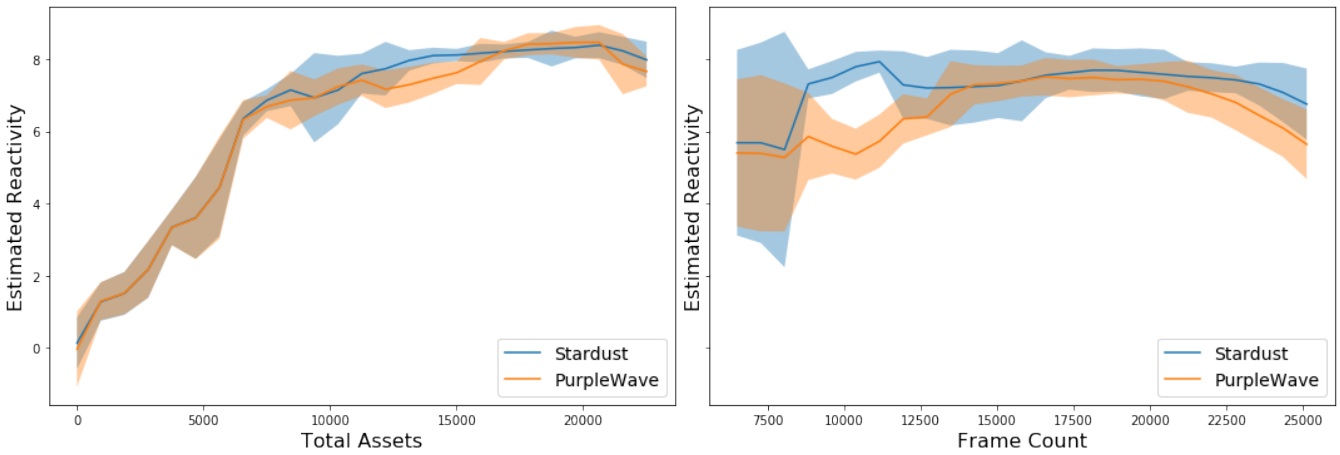

Figure 3: Stardust (Blue), PurpleWave (Orange). Left: Top agents indicate nearly identical reactivity across time. Right: First-place Stardust shows slightly better reactivity in the lower asset levels relative to second-place PurpleWave.

in magnitude between the air reactivity and cloak reactivity is extremely small relative to the standard errors, suggesting approximately equal levels of reactivity. We attribute this distinction to CRF's ability to search for interaction between individuals and game factors such as army size or game time, flexiblity that DiD does not provide.

In columns 1 and 2, we instead use L_{itg} to estimate anticipation effects. However, CRF did not converge, indicated by the large negative coefficients and untenable standard errors. This did not appear as a problem for DiD. We attribute this lack of convergence to a violation of α-regularity; there is not sufficient dispersal of L_{itg} periods - air risks tend to be first observed at the same time every game, and there is (at most) a single period before air arrives. As a result, the propensity of rare events tend to be extremely close 1 or 0. Since these propensities are placed in the denominator, the standard errors were nearly ± 10000 or failed to be measured. In other packages (e.g. *grf* for R), there are standard error messages for this case.

We next discuss examples of individual agents, rather than the aggregate average agent, and use the CRF method to explore and characterize reactivity along different features.

Discussion

We have identified (Table 2 and Table 3) the average agent in the 2020 competition significantly responds to the appearance of cloak and air risks. This identification stems specifically from the fact that frames following the discovery of risk have substantially more units that mitigate that risk relative to the agent's average for that time and type of game. The baseline created by the extensive fixed effects allows us to control for many cases, such as when agents are simply following a regular build order and coincidentally building protective units at the same time risks are observed. Instead, we have evidence that the average agent is reacting to the information within the game that is simultaneous with the appearance of the threat. Individually, we find a great deal of variation in agent performance, see Figure 5, placed in the appendix for reference. A detailed version of any agent's figure is available upon request or can be manufactured with the code and data found online at [Author's Github]. Despite this variation, nearly every single agent has significant and positive reactions to air risks, at nearly all frame counts. These reactions tend to improve as the agents gather more assets - catching the fact that agents are more willing and able to respond when they have more resources to do so. Both of the top two agents as measured by winning percentage (Stardust and Purplewave) indicate nearly identical response charts across time and across mitigation levels, expanded and overlayed in Figure 3.

In Figure 3, estimates of their reactivity under these counterfactual situations are always positive and significant, with the necessary exception of when they have no units. Particularly interesting is the shape of McRave's nonreaction in early periods, a capable agent that is now a regular finalist. We have expanded this plot of reactivity in Figure 4. One might hypothesize that at the competitive level, all agents react at all time periods, particularly for McRave's agent, which has risen in the rankings rapidly in the last few years. An interview with McRave's author suggests that this is indeed reasonable. Discussion reveals that the 2020 version of the agent did not react to air risks in the early game - it did not build Spore Colonies or Hydralisks to defend. However, McRave's author adds that in very late-game, the agent transitions to an army that is allowed to include Hydralisks, matching the frame counts where we see the increase in reactivity. We see direct confirmation as a critical sanity check to confirm the feasibility of these results.

Conclusion

Ability-based evaluations are essential to moving past single task-based evaluations and gaining a more nuanced understanding of AI systems. Risk management is a key ability in everyday life tasks and warrants methods to evaluate the abilities of AI systems. To this end, we have made three contributions that both further ability-based evaluations and our

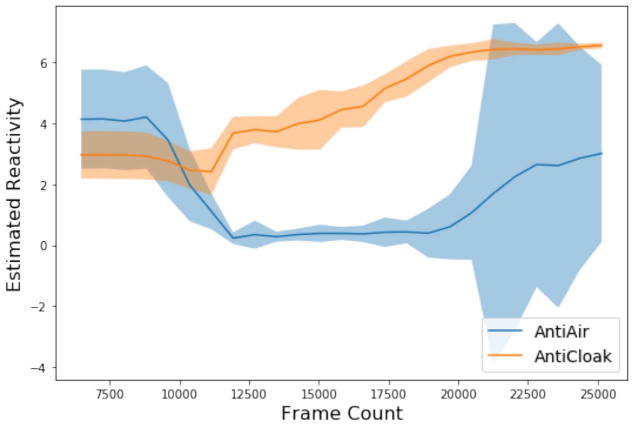

Figure 4: McRave's anti-air reactivity (blue) is nearly zero for much of the mid-game until very late game while anti-cloak reactivity (orange) is steadily increasing over time.

knowledge of agent risk management.

Our first contribution is the definitions of two risk management strategies: anticipation and reaction. While our definitions are designed for StarCraft, they can be easily adapted to other domains, supporting cross-domain ability-based evaluations. Second, we performed a risk management ability-based evaluation of the agents in a recent StarCraft AI Tournament using two different causal inference techniques. We find that many agents react to observing risks by investing in mitigations. We also found that the average tournament-level agent anticipates risk by investing in mitigations before risks are even observed. Finally, we provide the data and tools we used to extract relevant StarCraft data. This enables others to investigate more domains, reinforcing the qualities that make risk management a worthwhile ability-based evaluation.

Our most immediate future work tasks consists of analyzing additional years of StarCraft tournaments to identify longitudinal patterns of risk management. This will both expose any limitations of our current results (e.g. a confounding variable in 2020) and inform a more general way to assess how agents manage risk (e.g. beyond hard counters). More ambitiously, we'd like to develop the ability to assess agents outside of gameplay, as the number of games needed for the analysis we presented here is onerous, and replace it with a pre-hoc risk management assessment using a more general concept such as the Anticipatory Thinking Assessment Framework (Amos-Binks, Dannenhauer, and Gilpin 2023).

CRF

Here, we move away from aggregated measures and instead look at the estimated individual reactions for each agent, displayed as sparklines along potential inputs. We cannot examine individual anticipations for agents because of the same problem mentioned in Table 3, the propensities of treatment for these narrow timing attacks approach 0/1 and the results become unstable.

Figure 5 reinforces the intuition of Table 3. All agents in Column 1 show low reactivity when they have a low total market value. This is in line with expectations - an agent with no assets cannot react, and so all agents begin this column at the origin. Most agents are generally more reactive as their asset count increases, though some agents decline in reactivity as their assets grow larger - perhaps something indicative of fixed late-game composition or naturally building into anti-air units (such as Goliaths or Carriers).

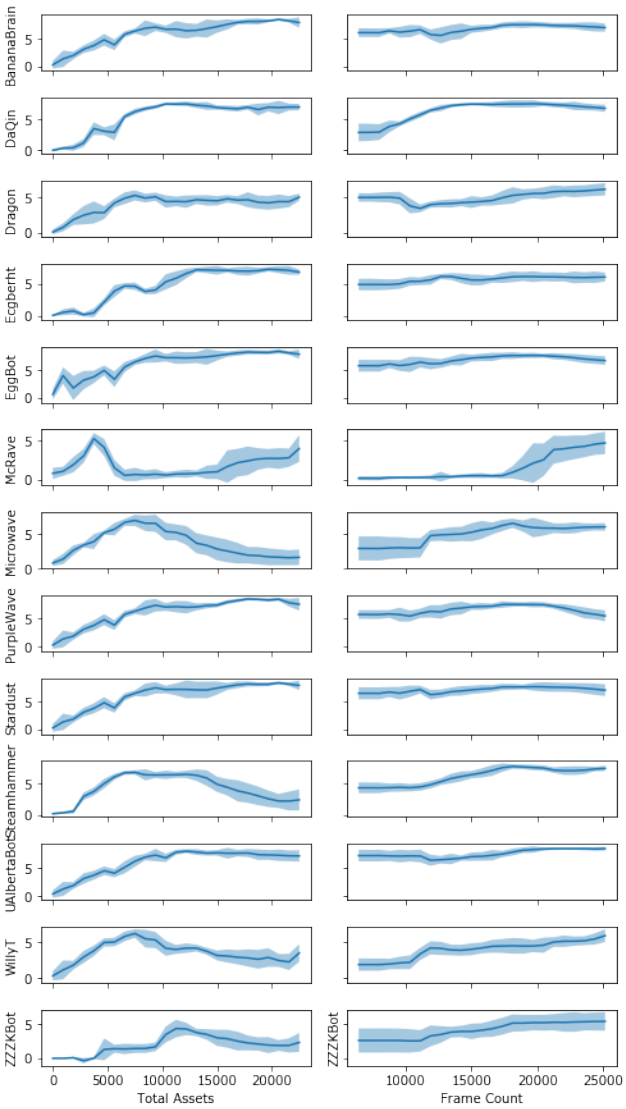

Figure 5: The blue line represents the predicted reactivity of the agent at a given total market value or frame count. This reactivity is estimated as their preferred race (random was represented as 1/3 of all other races), and at the mean of all other features. The shaded blue area is the 95% confidence interval.

In Column 2, we plot how an agent's reactivity changes over time. Reactivity tends to increase with the game duration until a plateau is reached. Some agents contain regions where their reactivity declines (holding other factors

Air	Risks ($\mathcal{R}^{air}$)	Mitigations ($\mathcal{M}^{air}$)
Protoss	Carrier, Interceptor, Scout, Arbiter, Corsair	Scout, Corsair, Dragoon, Arbiter, Archon, Interceptor, Photon Cannon
Terran	Science Vessel, Battlecruiser, Wraith	Battlecruiser, Ghost, Marine, Wraith, Valkyrie, Missile Turret, Goliath
Zerg	Mutalisk, Cocoon, Queen, Guardian	Devourer, Mutalisk, Hydralisk, Spore Colony, Scourge
Cloak	Risks ($\mathcal{R}^{cloak}$)	Mitigations ($\mathcal{M}^{cloak}$)
Protoss	Observer, Dark Templar	Observer, Photon Cannon
Terran	Wraith, Ghost	Comsat Station, Missile Turret, Science Vessel, Vulture Spider Mine
Zerg	Lurker	Overlord, Spore Colony

Table 4: Air and cloak units pose risks and serve as mitigations but are not evenly distributed across races; we control for race in our analyses.

at the mean), indicating potential opportunities to examine for weaknesses.

Unit Capabilities

StarCraft units that we used to calculate air and cloak risks/mitigations are in Table 4.

Acknowledgements

We would would like to thank the many thoughtful comments by reviewers and the insightful discussions with colleagues that helped shape and influence this article. This work is supported by DARPA contract HR001121C0030.The views expressed in this paper are those of the authors and do not necessarily represent the official position or policy of the U.S. Government, or the Department of Defense.

References

Amos-Binks, A.; Dannenhauer, D.; and Gilpin, L. H. 2023. The anticipatory paradigm. *AI Magazine*.

Angrist, J. D.; and Pischke, J.-S. 2008. Parallel worlds: fixed effects, differences-in-differences, and panel data. In *Mostly harmless econometrics*, 221–248. Princeton University Press.

Athey, S.; Tibshirani, J.; and Wager, S. 2019. Generalized random forests. *The Annals of Statistics*, 47(2): 1148–1178.

Athey, S.; and Wager, S. 2019. Estimating treatment effects with causal forests: An application. *arXiv preprint arXiv:1902.07409*.

Balazsi, L.; Matyas, L.; and Wansbeek, T. 2017. Fixed effects models. *The Econometrics of Multi-Dimensional Panels: Theory and Applications*, 1–34.

Ballinger, C.; and Louis, S. 2013. Comparing Coevolution, Genetic Algorithms, and Hill-Climbers for Finding Real-Time Strategy Game Plans. In *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '13 Companion, 47–48. New York, NY, USA: Association for Computing Machinery. ISBN 9781450319645.

Buro, M.; Churchill, D.; Churchill, D.; and Churchill, D. N. 2012. Real-Time Strategy Game Competitions. *AI Magazine*.

Callaway, B.; and Sant'Anna, P. H. 2021. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2): 200–230.

Card, D.; and Krueger, A. B. 1994. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4): 772–793.

Čertickỳ, M.; and Churchill, D. 2017. The current state of StarCraft AI competitions and bots. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 13.

Čertickỳ, M.; Churchill, D.; Kim, K.-J.; Čertickỳ, M.; and Kelly, R. 2018. StarCraft AI Competitions, Bots and Tournament Manager Software. *IEEE Transactions on Games*, PP: 1–1.

Churchill, D.; Buro, M.; and Kelly, R. 2019. Robust continuous build-order optimization in StarCraft. *IEEE Conference on Computatonal Intelligence and Games, CIG*, 2019-Augus.

Churchill, D.; Churchill, D.; Churchill, D. N.; and Buro, M. 2011. Build order optimization in StarCraft. *Artificial Intelligence and Interactive Digital Entertainment Conference*.

Geden, M.; Smith, A.; Campbell, J.; Spain, R.; Amos-Binks, A.; Mott, B.; Feng, J.; and Lester, J. 2019. Construction and Validation of an Anticipatory Thinking Assessment. *Frontiers in Psychology*, 10(December): 1–10.

Gunning, D.; Vorm, E.; Wang, J. Y.; and Turek, M. 2021. DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*, 2(4): e61.

Heinermann, A.; et al. 2012. BWAPI—An API for interacting with StarCraft: BroodWar. *available from https://github.com/bwapi/bwapi (accessed 2014-02-03)*.

Hernández-Orallo, J. 2017. Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. *Artificial Intelligence Review*, 48(3): 397–447.

Hernández-Orallo, J.; Baroni, M.; Bieger, J.; Chmait, N.; Dowe, D. L.; Hofmann, K.; Martínez-Plumed, F.; Strannegård, C.; and Thórissons, K. R. 2017. A new AI evaluation cosmos: Ready to play the game? *AI Magazine*, 38(3): 66–69.

Hirano, K.; Imbens, G. W.; and Ridder, G. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4): 1161–1189.

Levesque, H.; Davis, E.; and Morgenstern, L. 2012. The winograd schema challenge. In *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*.

Ontañón, S.; Synnaeve, G.; Uriarte, A.; Richoux, F.; Churchill, D.; Churchill, D.; Churchill, D. N.; and Preuss, M. 2013. A Survey of Real-Time Strategy Game AI Research and Competition in StarCraft. *IEEE Transactions on Computational Intelligence and AI in Games*.

Rosenbaum, P. R.; and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41–55.

Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. 2017. Mastering the game of go without human knowledge. *Nature*, 550(7676): 354–359.

Stock, J. H.; and Watson, M. W. 2012. *Introduction to econometrics*, volume 3. Pearson New York.

Synnaeve, G.; and Bessiere, P. 2012. Special tactics: A Bayesian approach to tactical decision-making. In *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, 409–416. IEEE.

Vinyals, O.; Babuschkin, I.; Czarnecki, W. M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D. H.; Powell, R.; Ewalds, T.; Georgiev, P.; et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782): 350–354.

Wager, S.; and Athey, S. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

Weber, B. S. 2018. Standard Economic Models in Nonstandard Settings – StarCraft: Brood War. *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 1–8.

Weber, B. S.; and Cappellari, P. 2022. Assessing the Impact of Ferry Transit on Urban Crime. *Urban Affairs Review*, 1–23.