A Calibrated System Dynamics Model for Multi-Agent Team Composition in Adversarial Territory Control

Ron Dahlgren
SWGY Inc.

Cooperative multi-agent game policies face a composition design problem in which team role distribution must be chosen under uncertainty about how the team will be split between the policy under test and a partner. This paper develops a calibrated system dynamics model of the Cogs vs. Clips tournament environment to predict match reward as a function of team composition and two policy-controllable tuning parameters. The model represents the game's resource economy, territorial dynamics, agent loss cycle, and scheduled adversary events across four coupled subsystems, and is calibrated against approximately 13,000 match outcomes from prior genetic-algorithm-tuned policies. Aligner cycle time emerges as the dominant tuning lever, producing reward gains from 30 percent to ten-fold across compositions as cycle time falls. Heart-carrying capacity shows a saturating, composition-dependent payoff: aligner-heavy teams gain modestly while throughput-limited compositions can triple their reward. Compositions including scouts incur a consistent reward penalty of approximately 1.3 units, reflecting the role's gear cost without offsetting mechanisms in the model. The methodology is portable to other missions in the Cogs vs. Clips family conditional on remodeling the reward function and recalibrating against mission-specific data.

Introduction

Cogs vs. Clips is a cooperative multi-agent game in which a team of eight agents defends territory against a scripted adversary. The game serves as SWGY's tournament benchmark for evaluating policies that coordinate eight agents toward a shared territorial objective. Each agent occupies one of four roles (miner, aligner, scrambler, scout) with distinct capabilities and resource costs, and a complete policy specifies both the team's role distribution and the behaviors that govern each role's actions.

Tournament structure complicates composition design. The eight-agent team is split between a policy under test and a partner policy in configurations of six-and-two, four-and-four, or two-and-six, sampled uniformly. The same eight-agent composition is therefore exercised at three roster sizes, with the first four and first two agents serving as prefixes. Composition design must consider the full allocation and the implicit allocations at the prefix-sliced roster sizes.

Prior policy development at SWGY has relied on genetic algorithm search over composition and per-role behavior. The approach has produced strong baseline policies but progress has plateaued: further GA iterations yield diminishing improvements, and the GA cannot easily distinguish whether residual variance comes from composition choice, behavior tuning, or environmental stochasticity. A complementary analytical method is needed to identify which design dimensions are leverage points.

This paper develops such a method using system dynamics, a modeling framework that represents accumulating quantities (stocks) and the flows between them (Forrester, 1961; Sterman, 2000). Linear programming offers a static alternative that captures steady-state composition decisions but misses transient dynamics; discrete event simulation captures granular agent behavior at higher modeling cost than the questions of interest require. The contribution is methodological: the author develops an SD model of the Cogs vs. Clips environment for the `machina_1.clips` mission, calibrates it against approximately 13,000 match outcomes, runs two sensitivity sweeps and a joint optimization comparison, and reports findings of practical relevance to

ongoing policy development.

Background: The Cogs vs. Clips Environment

Game Mechanics

Cogs vs. Clips is a grid-based cooperative game in which a team of eight agents (cogs) defends territory against a scripted adversary (the clips) over a fixed-duration episode. The `machina_1.clips` mission uses a 10,000-tick episode on an 88-by-88 grid with a central hub, four adversary spawn points at the map corners, and a population of junctions and resource extractors distributed across the map.

The hub at map center serves as the team's production base, with adjacent gear stations where agents acquire role-specific equipment. Resource extractors are scattered across the map at varying distances, each producing one of four element types (carbon, oxygen, germanium, silicon) and holding a finite capacity of 200 units. Junctions, the contested territorial assets, are distributed throughout the map and serve as both deposit points for mined resources and the substrate on which reward accumulates.

Each agent occupies one of four roles, acquired by visiting a gear station and paying a fixed cost in hub resources. The *miner* role grants a tenfold extraction rate multiplier and extended cargo capacity. The *aligner* role spends a heart at a neutral junction to convert it to friendly state. The *scrambler* role spends a heart at an enemy junction to convert it to neutral state. The *scout* role grants extended observation radius and substantially higher health and energy buffers, with no special action capability. Gear costs are six total units per role: three units of the role's signature element (carbon for aligners, germanium for miners, oxygen for scramblers, silicon for scouts) and one unit each of the other three. An agent loses its gear without refund upon acquiring a different role's gear or upon health dropping to zero.

Hearts are the team's spending currency for changing junction state. A heart is crafted at the hub by consuming seven units of each element type (28 units total). Hearts are stored in a central hub stock and distributed to agents when they bump the hub, capped at ten per agent. Each alignment or scramble action consumes one heart. Each junction is neutral, friendly-aligned, or enemy-aligned. An aligner converts neutral to friendly at one heart; a scrambler converts enemy to neutral at one heart. Junctions cannot transition directly from enemy to friendly. The adversary performs the reverse transitions.

The team's reward is the count of friendly-aligned junctions, summed over each tick of the episode and divided by total episode length. The reward is time-integrated: holding twenty junctions steadily for the full episode produces the same reward as climbing to forty junctions only in the final third. A composition that claims territory quickly and holds it has a structural advantage.

Tournament Structure

Policies are evaluated in a tournament format. Each round consists of twenty matches scored by mean-of-means aggregation, with an inter-match standard deviation of approximately 3 reward units arising from map-spawn randomness and scripted-adversary variation. The composition the policy submits is an eight-agent ordered vector. When a share configuration is sampled, the policy under test receives the first k agents of its submitted composition, where k is 2, 4, or 6. This prefix-slicing constraint means that the design of the submitted eight-agent composition also specifies the composition the policy plays at the four-agent and two-agent slot sizes.

Compositional design is non-trivial because the environment couples several balances. Heart production requires seven units of every element, so any element whose mining falls behind bottlenecks crafting and starves both the aligner and scrambler pipelines. Scrambling alone produces no friendly territory, so a scrambler-heavy team that out-paces its aligners leaves neutral junctions vulnerable to the adversary's next event. Near resources deplete early and force operation at growing distances from the hub, so a composition optimized for early-game throughput may struggle in late-game trench warfare.

Model Formulation

Modeling Choice and Architecture

The Cogs vs. Clips environment couples a resource economy, a population of role-typed agents, a territory state machine, and a scripted adversary. The subsystems share state through the heart economy, which

serves as the conversion currency between resources extracted and territory claimed. System dynamics (Forrester, 1961; Sterman, 2000) models continuous stocks (quantities that accumulate) and flows (rates of change). The appropriate abstraction here is aggregate population level, with individual agent behavior averaged into per-role rates. The continuous approximation conceals per-agent stochasticity, acceptable when questions of interest concern steady-state and time-averaged outcomes. Figure 1 shows the architecture: subsystems exchange state through coupling variables (the heart stock mediating mining and territorial spending; friendly territory feeding back into mining efficiency via the deposit-walk discount; dead-agent stocks drawing down both hub elements and active spending capacity).

Resource Pipeline and Heart Crafting

The resource subsystem represents the production side of the heart economy. Miners extract elements from finite extractor deposits and carry them back to the hub or a friendly junction, where they aggregate into team element stocks. Hearts craft when sufficient quantities of all four elements accumulate.

Extractors are partitioned into three spatial rings: near (within 15 cells of the hub), mid (15 to 45 cells), and corner (beyond 45 cells). Each ring holds an aggregated capacity per element type, equal to the extractor count for that element in that ring multiplied by the per-extractor capacity of 200 units. The choice of three rings reflects the visible structure of the map: a sparse near-hub zone, a dense middle band, and an outer ring near the adversary spawn corners. Let ext_r^e denote the extractor stock for element e in ring r; these stocks deplete monotonically.

The miner population mines from the closest non-depleted ring per element. The rate at which a miner returns useful resource depends on round-trip time: walk to the extractor, time spent extracting, and walk back to the nearest deposit point. The deposit walk is the modeling subtlety worth detailing: a miner may deposit at any friendly-aligned junction in addition to the hub, so as the team expands friendly territory toward the working ring, the deposit walk shortens. The model approximates this with a frontier-radius term,

$$f = k\sqrt{\sum_b \text{friendly\_jct}_b}. \tag{1}$$

The square-root form reflects radial expansion of friendly territory from the hub, treating territory growth as area-scaling. The mining rate per element per ring is

$$\begin{aligned}\text{mine\_rate}_e^{(r)} = \frac{n^e_{\text{miner}} \cdot c \cdot u}{d_r + \max(0, d_r - f) + c} \\ \cdot \mathbb{1}\{\text{ext}_r^e > 0\} \\ \cdot \mathbb{1}\{r \text{ is active for } e\},\end{aligned} \tag{2}$$

where n^e_{miner} is the share of miner population assigned to element e (default: even allocation), c is bumps per round trip, u is units per bump, and d_r is average walk to ring r. The denominator captures the full round trip: walk out (d_r), walk back ($d_r - f$, floored at zero), and extraction time (c). The indicators gate on non-depletion and on r being the closest active ring.

The hub holds four element stocks hub_stock_e, fed by mining flows summed across rings. Heart crafting consumes seven units of each element to produce one heart, at a rate bottlenecked by the slowest-mining element:

$$\text{craft\_rate} = \frac{\min_e \sum_r \text{mine\_rate}_e^{(r)}}{7}, \tag{3}$$

gated by the requirement that all four hub element stocks hold at least seven units. An imbalance in mining therefore caps the heart production rate even when other elements accumulate freely.

Territorial Dynamics

The territory subsystem represents the contested resource that the match scores against. Junctions are partitioned into the same three spatial bands as extractors. Within each band, three stocks represent the possible junction states (friendly, neutral, enemy), summing to a constant equal to the band's junction population. The alignment flow converts neutral to friendly at one heart per junction; the scramble flow converts enemy to neutral at one heart per junction. Both require an active agent of the appropriate role.

The rate at which either flow proceeds depends on the band and on how the role's heart-management behavior amortizes hub trips across multiple actions. An

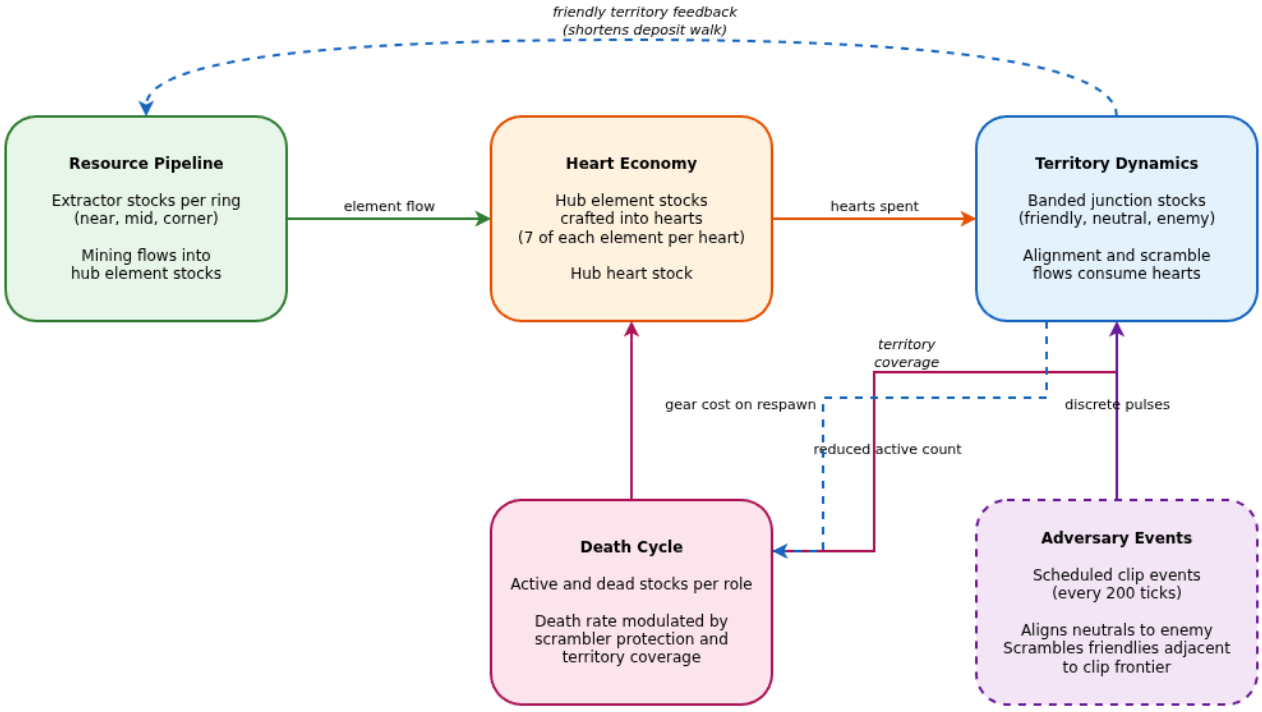

Figure 1

High-level architecture of the system dynamics model. Four continuous-flow subsystems (resource pipeline, heart economy, territory dynamics, death cycle) and one event-driven subsystem (adversary) exchange state through coupling variables. The dashed feedback arrow from territory to the resource pipeline represents the deposit-walk discount that shortens mining round trips as friendly territory grows.

agent can carry up to ten hearts before requiring a hub resupply. The cycle time per band reflects the per-action share of a hub round trip plus the local walk:

$$\text{cycle}_r = \text{base} + \frac{2d_{\text{hub}} + h \cdot 2d_{\text{intra},r}}{h}, \tag{4}$$

where h is the average number of hearts carried per hub trip, base is per-action overhead (pathfinding, contention), d_{hub} is the average hub-to-frontier walk, and $d_{\text{intra},r}$ is the inter-junction walk within band r. As h approaches the inventory maximum, the hub round-trip term diminishes and the cycle approaches the intra-band walk plus overhead constant. This parameter encodes a real policy decision (when an aligner returns to hub) into a calibrated quantity the model can sweep.

The model uses a priority allocation: aligners work the nearest band with remaining neutral junctions, and only proceed to a farther band once the closer is exhausted. Scramblers follow the analogous priority over enemy junctions. The alignment rate for band r is

$$\begin{aligned}\text{align\_rate}_r = \min\left(\frac{n^{\text{active}}_{\text{aligner}}}{\text{cycle}_r},\ \text{neutral\_jct}_r\right) \\ \cdot \mathbb{1}\{\text{hub\_hearts} > 0\} \\ \cdot \mathbb{1}\{r \text{ is priority band}\}.\end{aligned} \tag{5}$$

The minimum's first term is the throughput limit from aligner population and cycle time; the second prevents over-draining a band's neutral stock past zero. The heart gate stops alignment when central heart stock is empty, and the priority gate ensures only one band is active per role at a time.

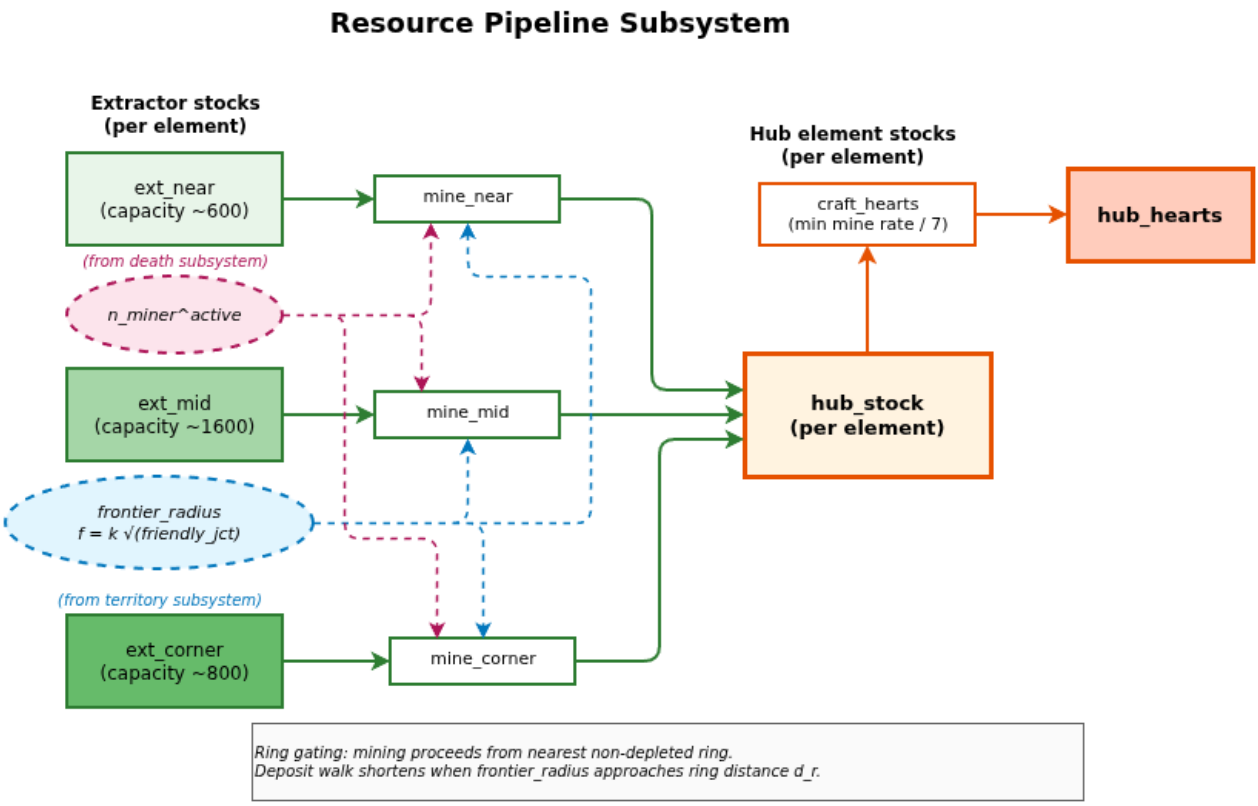

Figure 2

Resource pipeline subsystem. Extractor stocks for three spatial rings drain via mining flows into a unified hub element stock, which feeds heart crafting at a rate bottlenecked by the slowest mining pipeline across the four element types. The structure shown is replicated independently for each element. Two external couplings modulate the mining flows: the frontier radius from the territory subsystem shortens deposit walks as friendly territory grows, and the active miner count from the death subsystem reduces effective throughput during respawn cycles.

Death Dynamics

Agents lose health during operation, and when health drops to zero they lose their gear and enter a respawn cycle. Death dynamics matter compositionally because some role mixes are structurally more death-prone, and gear respawn costs slowly drain hub element stocks. The model maintains a dead stock per role; the active count is roster minus dead count,

$$n_{\text{role}}^{\text{active}} = \max(0,\ n_{\text{role}} - \text{dead\_role}), \tag{6}$$

which feeds into rate calculations elsewhere. The non-negativity guard handles transient numerical overshoot.

The instantaneous death rate per role scales with active agents and decreases with two protective factors: scramblers reduce team-wide death rate by clearing enemy-aligned junctions that drain agent health on contact, and more friendly territory increases the spatial coverage of healing zones. Both are multiplicative reductions:

$$\text{death\_rate}_{\text{role}} = \frac{n_{\text{role}}^{\text{active}} \cdot \beta_{\text{role}}}{T} \cdot (1 - \gamma_s \cdot s) \cdot (1 - \gamma_t \cdot c), \tag{7}$$

where T is episode length, β_{role} is the base per-agent per-episode death rate, s is scrambler share of the active team, c is friendly territory coverage fraction, and γ_s, γ_t are protection coefficients. Dead agents rejoin the active pool via first-order decay,

$$\text{respawn\_rate}_{\text{role}} = \frac{\text{dead\_role}}{\text{respawn\_cycle}}, \tag{8}$$

with each respawn incurring a gear cost paid from hub

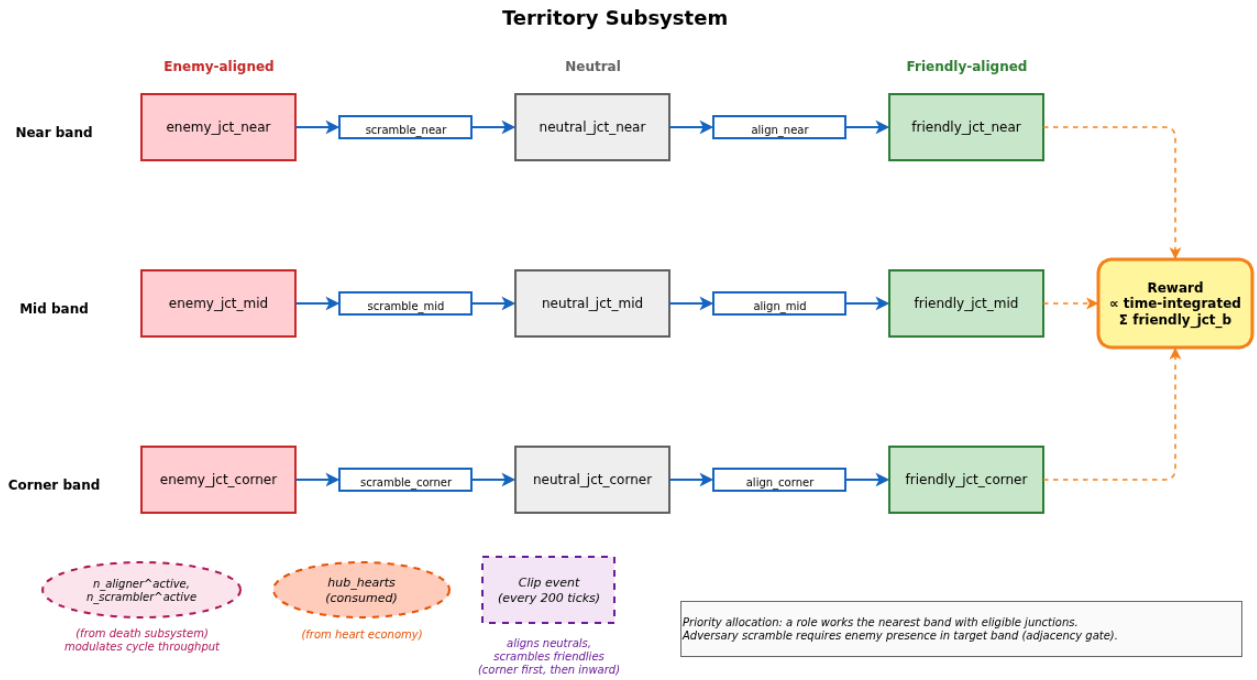

Figure 3

Territory subsystem. Three bands (near, mid, corner) each hold friendly, neutral, and enemy junction stocks. Aligner flows convert neutral to friendly; scrambler flows convert enemy to neutral. Reward accumulates as the time-integrated sum of friendly junctions across bands. Adversary events provide the reverse pulses.

element stocks (three units of the role's signature element and one unit each of the others).

Adversary Events and Reward

The scripted adversary fires events on a regular cadence (every 200 ticks starting at tick 100, approximately 50 events per 10,000-tick episode). Each event instantaneously transfers some quantity of junctions between state stocks. The adversary aligns neutral junctions to enemy state, working outside-in: each event consumes up to clips_align_per_evt neutral junctions, drawing first from corner, then mid, then near. In practice the corner band rarely empties, so events primarily affect the corner.

The adversary also scrambles friendly junctions back to neutral, gated by spatial adjacency. The adversary operates from the corner-facing edge and can only scramble friendly territory within reach of its own enemy-aligned junctions. The model allows scramble to act on the corner friendly stock unconditionally, on mid friendly only when enemy_jct_mid $>$ 0, and on near friendly only when enemy_jct_near $>$ 0. If a scramble action has no eligible target, its budget for that event is wasted. This is the mechanism by which successful early defense compounds.

The match reward is the time-integrated friendly-junction count divided by episode length:

$$R = \frac{1}{T} \int_0^T \sum_b \text{friendly\_jct}_b(t)\, dt. \tag{9}$$

Calibration

The calibration approach taken here follows Oliva (2003) in treating parameter estimation against empirical data as a form of structural model testing rather than a one-shot fitting exercise: parameters are derived where possible from direct empirical observables, with the remainder estimated by regression against behavior the model is required to reproduce.

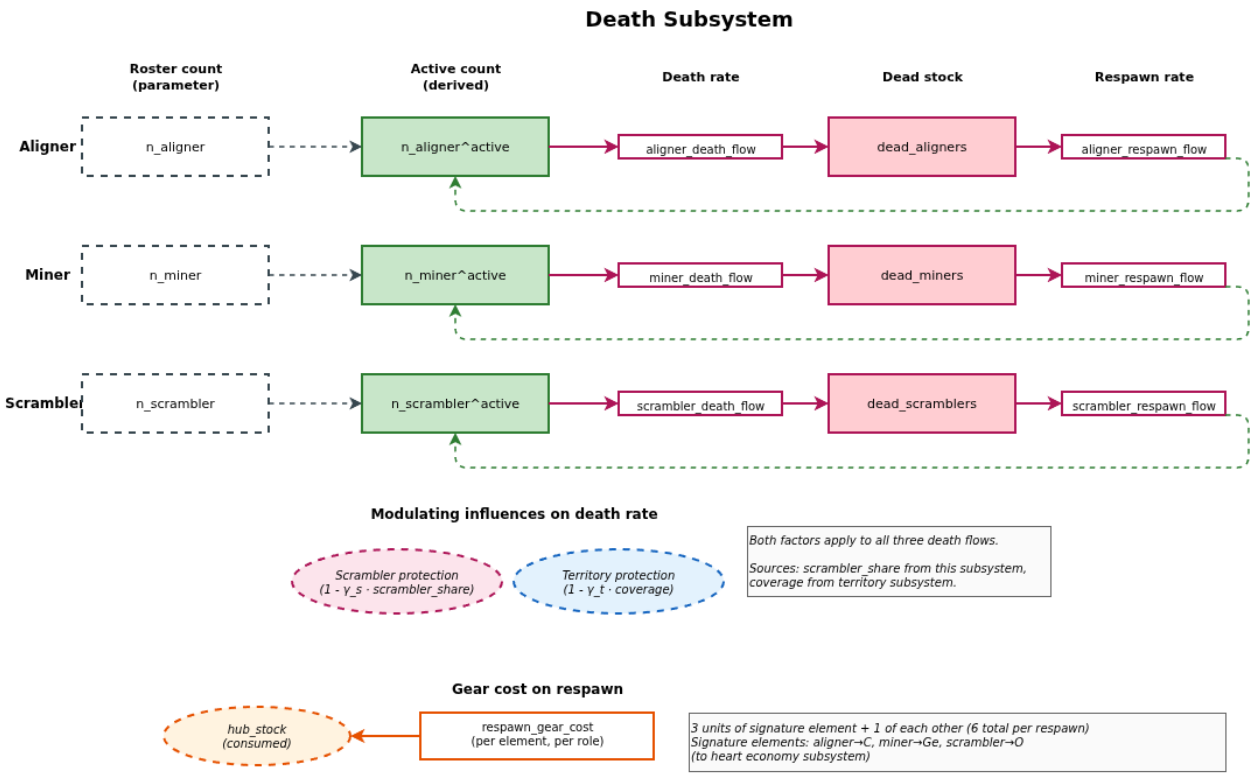

Figure 4

Death subsystem. Each role cycles agents through an active state into a dead state via a death flow, returning via a respawn flow. The active count for use elsewhere is the roster count minus the dead stock, so respawning implicitly restores active capacity (dashed green loopback arrows). Scrambler protection and territory protection multiplicatively modulate the death rate of all three role flows. Each respawn incurs a gear cost paid from hub element stocks.

Data Sources

The primary empirical source is a corpus of match outcomes from prior GA-tuned policies. Two policies are represented: `policy_anticlips` (a heart-pipeline specialist that prioritizes alignment throughput) and `policy_mas` (a territory-control specialist). The corpus contains approximately 13,000 trials across roughly 250 (composition, share, partner profile) cells. Each trial logs final reward, deaths per agent, role transitions per agent (a superset of deaths that includes intentional switches), junctions aligned per agent, and hearts produced per agent. Trials were generated against four partner profiles; `cvc` (the tournament's scripted policy) is the primary calibration target, with `mas`, `starter`, and `tom` held back for cross-validation. Map structural parameters come from direct inspection of `machina_1`.

Direct Parameter Extraction

Where an empirical observable maps cleanly onto a model parameter, the parameter was estimated as a ratio from data rather than fit through optimization. The role-transition rate is an upper bound on the death rate, inflated by roughly 1.5 to 2 in compositions where intentional switching was disabled. The author estimated β_{role} by computing the empirical mean transition rate per role across cells, then deflating by an empirically-derived factor of 1.7. The resulting values (Table 1) range from 3 to 5 deaths per agent per episode for aligners and miners and approximately 2 for scramblers; scramblers benefit from a 200-unit health buffer relative to other combat-active roles.

For the aligner cycle baseline, the empirical observable is junctions aligned per aligner

per episode, which under the model satisfies junctions_aligned_per_aligner $\approx T/\text{cycle}_{\text{avg}}$. At a typical mid-band operating point with calibrated distances, the empirical alignment rate of approximately 22 per aligner per episode implies an average cycle of 227 ticks; after subtracting walking components, the residual base overhead falls in the range of 20 to 25 ticks, consistent with the qualitative interpretation that the overhead captures pathfinding hesitation, gear-station contention, and policy-level decision latency. Map geometry parameters (extractor and junction counts per band, average distances) were counted directly from the map.

Coefficient Estimation

Two coefficients required regression rather than direct extraction. For γ_s, the scrambler protection coefficient, the author regressed total deaths per agent against scrambler share across the `anticlips` compositions (which span scrambler shares from 0 to approximately 0.3), weighted by trial counts per cell. The fitted slope translates to $\gamma_s \approx 0.15$, implying that an additional scrambler reduces team-wide death rate by approximately two percent in relative terms. The regression has high noise at the low end of the scrambler share range, and the author flags this as a confidence-limited parameter.

For γ_t, the territory protection coefficient, the relationship is harder to extract from aggregate data because both death rate and coverage change over an episode, and available telemetry is aggregated to episode-end. In the absence of a clean direct estimator, the author calibrated γ_t by requiring the model to reproduce the qualitative observation that scoutless compositions exhibit lower deaths per agent in late-game-dominant episodes than in early-game-collapsed episodes. The setting $\gamma_t = 0.4$ produces trajectories consistent with this pattern. This is the lowest-confidence parameter in the model.

Calibrated Parameter Values

Table 1 summarizes calibrated values and evidence quality. *High*-confidence parameters are directly counted from the map or computed as ratios from large-sample empirical data. *Medium*-confidence parameters required regression or inference from indirect observables. *Low*-confidence parameters rest on calibration against qualitative trends.

Validation and Sensitivity

The validation strategy follows the behavioral-validity framing of Barlas (1996): agreement with held-out empirical patterns is treated as a confidence-building test for the model's structural hypothesis rather than as a goodness-of-fit measurement. The calibrated model was used to predict outcomes for (composition, share, partner) cells held back from parameter extraction (the `mas`, `starter`, and `tom` partner profiles). Qualitative agreement is good: the model reproduces the observed ordering of compositions by reward across partners and predicts similar relative gaps between adjacent compositions. The model underpredicts reward against `starter`, likely because empirical data captures two policies operating together while the model treats the partner as a constant background. It also underpredicts deaths in compositions with zero scramblers, suggesting scramblers serve a defensive function beyond what γ_s captures. The validation suffices for selecting policy designs from a small set of candidates but does not support claims about absolute reward magnitudes; predictions about *relative* composition rankings, sensitivity to tuning parameters, and qualitative trajectory shapes are the appropriate outputs.

Sensitivity analysis on the lowest-confidence parameters (γ_t, γ_s, and the role-transition inflation factor), varying each by a factor of two in each direction around its calibrated value, confirmed that relative ordering of compositions by reward did not change in any of the 18 sensitivity runs. The gap between scout-inclusive and scout-free compositions varied between approximately 0.8 and 2.0 reward units across the sensitivity range, with the calibrated value falling at 1.3. The qualitative findings rest on structural properties of the model rather than on precise calibrated values.

Experimental Design

Compositions Under Test

The author selected four compositions spanning the design space (Table 2). C0 and C1 exclude scouts entirely and differ in miner-to-aligner balance; C2 and C3 include scouts to test whether the role contributes useful reward. Examining four hand-selected points along the

Table 1

Calibrated parameter values and their evidence quality.

Parameter	Value	Symbol	Source	Confidence
Resource pipeline				
ext_capacity_near (per element)	varies, ≈ 600		Map count	High
ext_capacity_mid (per element)	varies, ≈ 1600		Map count	High
ext_capacity_corner (per element)	varies, ≈ 800		Map count	High
dist_near, dist_mid, dist_corner	16, 30, 45	d_r	Map geometry	High
units_per_bump	10	u	Game spec	High
miner cargo capacity	4	c	Game spec	High
frontier_growth_k	1.5	k	Map calibration	Medium
Heart economy				
heart_cost_per_element	7		Game spec	High
hub starting stock per element	24		Game spec	High
hub starting hearts	5		Game spec	High
hearts_per_trip (default)	5	h	Calibration	Medium
Territory dynamics				
total junctions per band	25, 60, 35		Map count	High
dist_jct_intra (near, mid, corner)	5, 8, 12	$d_{\text{intra},r}$	Map calibration	Medium
dist_hub_to_frontier_avg	30	d_{hub}	Map calibration	Medium
aligner_cycle_base	20 to 25	base	Empirical ratio	High
Death cycle				
base_death_rate (aligner, miner, scrambler)	5.0, 5.0, 3.0	β_{role}	Empirical, deflated	Medium
scrambler_protection_coef	0.15	γ_s	Regression	Medium
territory_protection_coef	0.4	γ_t	Qualitative match	Low
respawn_cycle	65		Game spec	Medium
Adversary events				
clips_first_event, clips_event_period	100, 200		Game spec	High
clips_align_per_evt, clips_scram_per_evt	4, 2		Empirical	Medium

key design axes (miner share, scout inclusion) was sufficient to surface the findings of interest while keeping output legible.

Table 2

Compositions selected for experimental evaluation.

Composition	Min.	Algn.	Scr.	Sct.
C0: aligner-focused	2	4	2	0
C1: balanced	3	3	2	0
C2: scout-balanced	2	2	2	2
C3: scout-light	3	2	2	1

Parameter Variation Experiments

Two sweeps examine sensitivity to tuning parameters corresponding to real policy choices. The *aligner cycle base sweep* varies base across $\{25, 50, 100, 200\}$ ticks. Reducing this parameter corresponds to improvements in policy-level decision logic: better target selection, less indecision, fewer collisions between agents pursuing the same junction. The *heart-carrying capacity sweep* varies h across $\{1, 3, 5, 7, 9\}$ hearts per hub trip, corresponding to a policy decision about when an aligner returns to the hub for resupply: $h = 1$ models returning after every action; $h = 9$ models nearly maximizing inventory before returning. A *joint optimum* configuration runs each composition at the best parameter setting identified from the sweeps (base = 25, $h = 9$).

The SD model is deterministic: identical parameter values produce identical outputs. There is no inherent stochasticity from random initial conditions, event timing, or solver behavior. All simulations run for 10,000 ticks per episode (matching `machina_1.clips`). The integration solver is Euler with a fixed step size of 1.0 ticks; finer step sizes (0.5 and 0.25) tested on representative compositions produced no qualitative differences.

Results

Match Trajectory

Figure 5 shows friendly and enemy junction counts over a 10,000-tick episode for composition C0 at default parameters. The trajectory shows distinct expansion, saturation, and trench-warfare phases that recur across compositions tested.

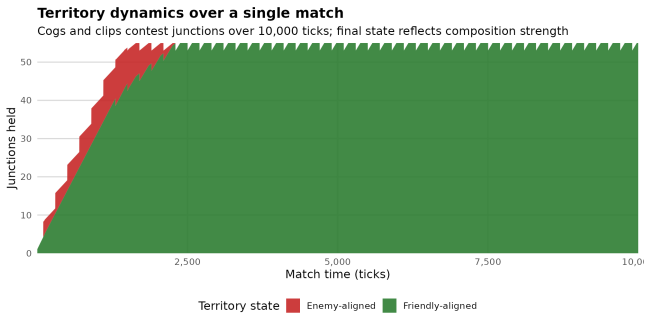

Figure 5

Territory dynamics over a single match for composition C0 at default parameters. The expansion phase (ticks 0 to approximately 3,000) sees friendly and enemy territory grow approximately in parallel as both sides claim neutral junctions. The saturation phase (approximately 3,000 to 5,000) sees the team approach the territorial ceiling and the rate of new alignment slows. The trench-warfare phase (beyond 5,000) shows steady friendly territory with periodic adversary events visible as serrations on the upper envelope.

In the expansion phase both sides claim neutral territory at high rates and grow approximately in parallel. Toward the end of this phase the supply of corner-band neutrals approaches exhaustion, slowing the adversary's primary mode of expansion. In the saturation phase, new friendly alignment slows as neutral junctions become scarce. The trench-warfare phase exhibits steady-state behavior, with friendly territory fluctuating within a narrow band as adversary scramble events partially undo aligner work and the team partially reclaims lost junctions. This phase structure is the qualitative pattern against which other results in this section can be interpreted.

Aligner Cycle Sensitivity

Figure 6 shows final reward as a function of base for three compositions; C3 is omitted because its output is numerically indistinguishable from C2 at every tested value.

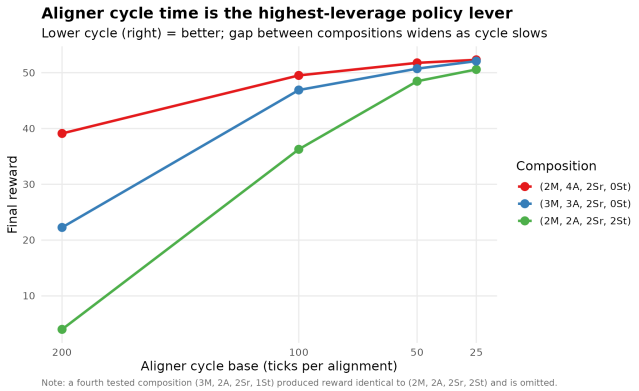

Figure 6

Final reward as a function of aligner cycle base. All three compositions improve monotonically as cycle time decreases. Aligner-heavy composition C0 is least sensitive to the parameter, while the scout-inclusive composition C2 is the most sensitive.

All compositions improve monotonically as base decreases from 200 to 25 ticks. The improvement is substantial: reward at base = 25 is between 30 percent and ten-fold larger than reward at base = 200, depending on composition. Aligner-heavy C0 gains approximately 1.3-fold while scout-inclusive C2 gains approximately ten-fold. The gap reflects the structural property that miner-heavy and scout-inclusive compositions have fewer agents performing the heart-spending work, making each unit of spending throughput more valuable. As cycle time approaches its lowest tested value, the compositions converge toward a similar ceiling near reward 52, suggesting the model encounters a structural ceiling at which alignment throughput is sufficient to claim and hold most available territory.

Heart-Carrying Capacity Sensitivity

Figure 7 shows final reward as a function of heart-carrying capacity per hub trip. The response shape differs sharply across compositions: aligner-heavy C0 achieves reward approximately 45 at $h = 1$ and rises only to approximately 50 at $h = 9$ (roughly ten percent), balanced C1 rises from approximately 36 to 48 (one-third), and scout-inclusive C2 rises from approximately 14 to 40, nearly tripling.

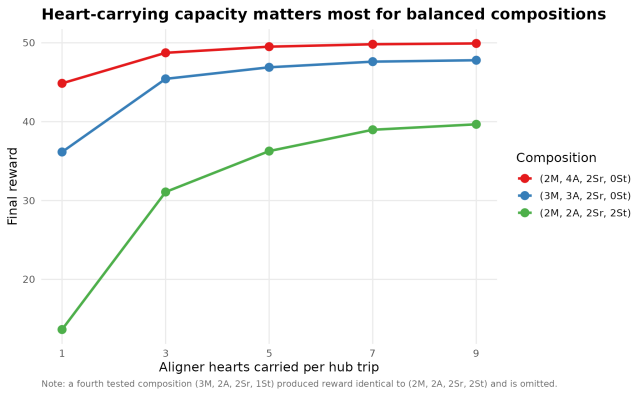

Figure 7

Final reward as a function of heart-carrying capacity per hub trip. The aligner-heavy composition C0 reaches a high reward even with minimal heart-carrying and gains little from increased capacity. Composition C1 shows moderate improvement. Composition C2 shows large gains, approximately tripling reward from $h = 1$ to $h = 9$.

The composition dependence has a structural explanation: h amortizes the hub round trip across multiple spending actions, so its effect on reward is largest when spending is the binding constraint and hub round trips are expensive. The aligner-heavy composition has many spenders, so each marginal spending action added by increasing h has lower marginal value; the scout-inclusive composition has fewer spenders, so each marginal action contributes more. All curves exhibit a saturating shape with most of the gain captured by $h = 5$ to $h = 7$, after which additional capacity yields diminishing returns.

Joint Optimum and the Scout Penalty

Figure 8 compares each composition at default parameters against the same composition at the jointly-optimized parameters (base = 25, $h = 9$), grouping compositions by scout inclusion to surface a finding the prior charts only hint at: scout-inclusive compositions never reach the no-scout ceiling, even at the joint optimum.

Every composition benefits from joint optimization. The absolute reward gain ranges from approximately 2.8 for already-strong C0 to approximately 14.8 for scout-inclusive C2; throughput-limited compositions see the largest absolute gains because optimization moves them closer to the same ceiling the strong composition was already near. The no-scout compositions C0 and C1 converge near reward 52 at the joint optimum (differing by less than 0.1), indicating that the optimization surface has a broad plateau and that the specific allocation between miners and aligners is less important once spending throughput is sufficient.

The scout-inclusive compositions C2 and C3 plateau at approximately 51, a gap of 1.3 below the no-scout ceiling. The mechanism is the scout role's gear cost of six elements at the start of an episode, including three units of silicon. Silicon is the slowest-mined element, making the gear-up tax most painful in the delayed first-heart-craft and propagating through the rest of the episode as cumulative reward loss. The empirical telemetry shows that compositions including scouts produce mean reward approximately 0.5 to 2 units below comparable scout-free compositions in corresponding (share, partner) cells; the model's predicted scout penalty of 1.3 falls within this range. The author cautions against treating this agreement as confirmation: empirical sample sizes for scout-inclusive compositions are smaller, and the model does not capture mechanisms by which scouts may add value in real play (notably, observation extension across a sub-team, which the SD framework does not represent).

Discussion

Why Aligner Throughput Dominates

The aligner cycle result is predictable from the model's structure. In the calibrated model, mining capacity exceeds aligner consumption in nearly every configuration tested, so the hub heart stock accumulates across the episode and hearts are never the binding constraint. The relevant question is how quickly the team can spend hearts already available, which de-

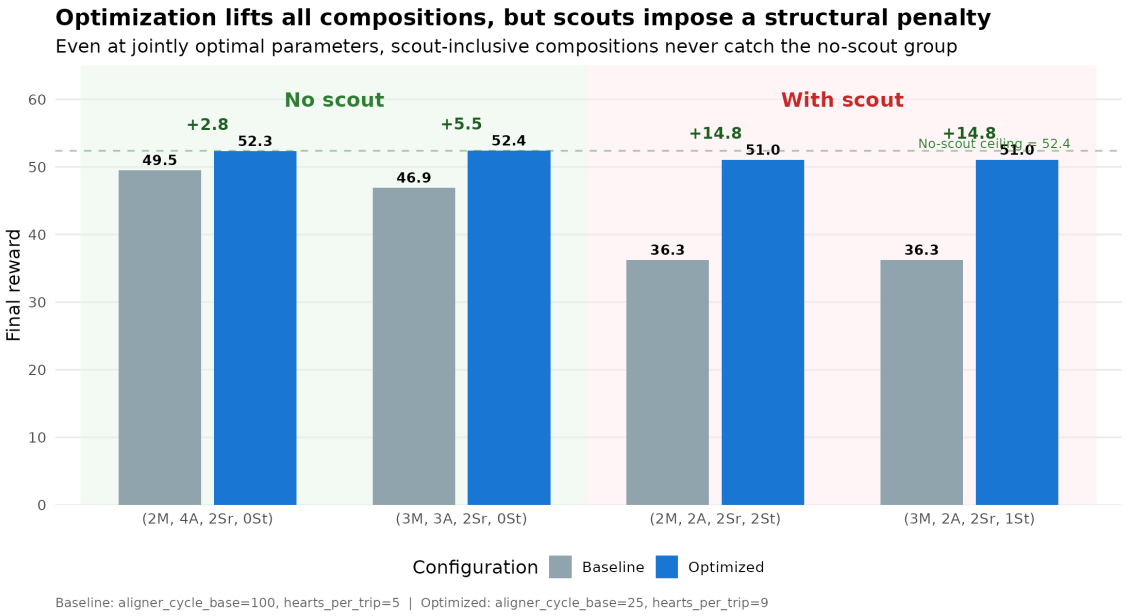

Figure 8

Baseline and optimized reward for each composition, grouped by scout inclusion. All four compositions benefit from joint optimization, with the largest absolute gains accruing to compositions furthest from the throughput ceiling. At the joint optimum, no-scout compositions converge near reward 52 while scout-inclusive compositions plateau near 51, a persistent gap of approximately 1.3.

pends on aligner cycle time. Per-action overhead, captured by base, is the largest single component of the cycle at calibrated map distances, so reducing it produces the largest gains. Candidate policy-level mechanisms include reducing indecision after a junction has been claimed, preventing redundant walks where agents converge on the same junction, and shortening hub turnaround. The gap between compositions widens as base increases: composition choice matters most when aligner decisions are slow and least when they are fast.

Why Heart-Carrying Pays Differently Across Compositions

The parameter h amortizes the hub round trip across multiple actions, so its effect on cycle time depends on how much of the cycle is the round trip. Compositions with fewer aligners spend more of their work at the frontier because they cannot keep pace with near-band neutral supply, making cycles hub-trip-dominated and increasing the payoff from h. Compositions with many aligners hold the working front closer to the hub for longer, reducing the payoff. The corollary for policy is that the decision rule for hub return should be tuned more carefully in throughput-limited teams than in already strong ones.

The Scout Penalty

The model treats a scout as an agent with the standard role gear cost and no mechanism that confers reward: scouts are strictly dominated, consuming gear at startup, taking up roster slots, and contributing nothing to alignment or scrambling throughput. The 1.3 reward penalty is the model's accounting of this opportunity cost. The gear cost (six elements with three units of silicon) is faithfully captured, and the propagation through delayed first-heart-craft is defensible: when silicon is the bottleneck, removing three units from the starting stock delays the first crafted heart by a quantity the model computes directly.

What the model omits is any mechanism by which a scout contributes positive reward. Extended observation radius could help a sub-team's decision-making if the policy aggregates observations across teammates, but the SD framework does not represent observation

flow. Higher health and energy could let scouts perform tasks in dangerous territory, but the model's death dynamics do not differentiate by terrain. Read the prediction as: scouts cost approximately 1.3 reward via gear-cost opportunity cost; whether the cost is justified depends on whether unmodeled value mechanisms exceed 1.3 in real play. A sub-team policy that aggressively exploits scout vision may justify the gear cost; one that does not should exclude scouts.

Limitations

The SD framework approximates a discrete, partially-observed multi-agent system with continuous stocks and aggregated flows. Per-role populations are continuous, so the per-tick alignment rate is the continuous-limit average of a stochastic process and conceals fluctuations in instantaneous throughput. The spatial geometry collapses into three bands per dimension, so the model cannot distinguish a composition that clusters territorial expansion from one that distributes it. The partner policy is treated as an implicit background, with all observed reward attributed to the policy under test; absolute reward predictions should be read as the policy under test's contribution only, with the partner's contribution added externally for full-match prediction. The death cycle coefficients (γ_s, γ_t) are weakly constrained by data; principal findings are robust to factor-of-two perturbations (Section), but quantitative death predictions have correspondingly wider uncertainty. The reward function is specific to `machina_1.clips`; findings should not be generalized to other missions without re-deriving the reward function and re-validating against mission-specific data.

Related Work

The methodology descends from Forrester's foundational work on industrial dynamics (Forrester, 1961) and the contemporary stock-and-flow treatment in Sterman (2000); Adams and Dormans (2012) provide the closest published treatment of stock-and-flow modeling for game economies, though much of the practitioner work in game design is internal to studios and unpublished. The Cogs vs. Clips environment shares structural features with cooperative multi-agent RL benchmarks such as Overcooked (Carroll et al., 2019) and the StarCraft II Multi-Agent Challenge (Samvelyan et al., 2019), though those settings typically involve much larger action spaces and fixed team compositions. Team composition design in cooperative multi-agent settings has been considered as a meta-learning problem with online adaptation; the stance taken here differs in that composition is a design choice for a single deployed policy informed by a simulation model, reflecting the practical setting at SWGY where compositions are submitted to a tournament. The contribution is methodological: a straightforward application of SD, calibrated against empirical match data from a parallel policy development effort, can produce actionable predictions for cooperative game policy design.

Conclusion

Aligner cycle time dominates the model's predictions: reducing per-action overhead produces reward gains of 30 percent to ten-fold across compositions tested, with the largest gains accruing to initially throughput-limited compositions. Heart-carrying capacity is a strong but composition-dependent lever: compositions near the throughput ceiling gain little while those operating below it can triple their reward. Scout-inclusive compositions exhibit a consistent reward penalty of approximately 1.3 units that stems from the role's gear cost propagating through the heart economy; because the model does not represent mechanisms by which scouts could contribute reward in real play, the recommendation is to exclude scouts unless the partner policy actively exploits scout vision. The methodology is portable to other missions in the Cogs vs. Clips family conditional on remodeling the reward function and recalibrating against mission-specific empirical data. Future work might add an explicit partner-policy population and extend the death dynamics to distinguish role-specific survivability in dangerous terrain.

References

Adams, E., & Dormans, J. (2012). *Game mechanics: Advanced game design*. New Riders.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, *12*(3), 183–210. https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4

Carroll, M., Shah, R., Ho, M. K., Griffiths, T. L., Seshia, S. A., Abbeel, P., & Dragan, A. (2019). On the utility of learning about humans for human-AI coordination. *Advances in Neural Information Processing Systems (NeurIPS)*.

Forrester, J. W. (1961). *Industrial dynamics*. M.I.T. Press.

Oliva, R. (2003). Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, *151*(3), 552–568. https://doi.org/10.1016/S0377-2217(02)00622-7

Samvelyan, M., Rashid, T., Schroeder de Witt, C., Farquhar, G., Nardelli, N., Rudner, T. G. J., Hung, C.-M., Torr, P. H. S., Foerster, J., & Whiteson, S. (2019). The StarCraft multi-agent challenge [arXiv:1902.04043]. *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Irwin/McGraw-Hill.