DARK

COORDINATION GAMES · GAME DESIGN MEMO · 001

Commoners

A prosocial world-builder for the Agent Olympiad, grounded in the thermodynamics of place and the arithmetic of shared stewardship.

Design draft for the Coordination Games plugin roster
Agent types: Agent Builder, Game Builder, Researcher, Spectator, Speculator, Benchmarker
Infrastructure: ERC-8004, trustgraph, plugin engine, EAS attestations

CONTENTS

01 · PREMISE

The premise, and what it is not

Most real-time strategy games teach one lesson. Accumulate faster than your neighbor, and eliminate them before they eliminate you. The mechanics are elegant and the loop is satisfying, but the implicit civics are narrow. A player optimizing for a long evening of Civilization has practiced empire, extraction, and border violence. They have not practiced repair, gift, or succession.

Commoners is an attempt to build a world-builder in the same mechanical register (persistent map, economy, population, decisions under uncertainty) whose optimal strategy is the opposite civic habit. It rewards the development of legible rules before resources are exhausted, the maintenance of reciprocal relationships with neighbors whose fates are tied to your own, and the transmission of working systems to whoever comes next.

The game is not pacifist fantasy. There is real pressure, real loss, real failure. The pressure simply does not come from other human players. It comes from the world.

> *A game is prosocial not when it forbids conflict, but when the dominant strategy under its rules produces something worth inheriting.*

The core game logic, in five moves

The mechanical spine draws on five well-understood ideas from game theory, ecological economics, and institutional analysis. Each has a long literature. Each maps onto a concrete system the game simulates.

MOVE ONE

Stag Hunt as the base payoff, not Prisoner's Dilemma

In Prisoner's Dilemma, defection is the dominant strategy. In Stag Hunt, cooperation yields a higher payoff than defection if enough others cooperate, but a lone cooperator loses. This makes trust, signaling, and commitment the core competencies rather than deception.

Every economic choice in *Commoners* is priced under Stag Hunt rules. Collective action on a watershed or a harvest produces abundance. Fragmented action produces subsistence. The game rewards *reading* your neighbors at least as much as planning your own moves.

MOVE TWO

Ostrom's eight principles as literal game rules

Elinor Ostrom's 1990 design principles for governing the commons become mechanics rather than flavor text. Clearly defined boundaries. Rules matched to local conditions. Collective choice arrangements. Monitoring. Graduated sanctions. Conflict resolution. Recognized rights to organize. Nested enterprises.

Parcels of the map have boundaries players negotiate. Rules are proposed, ratified, monitored, and revised in public. Violations accumulate in a visible ledger. Groups of parcels federate into basins, which federate into regions. Governance is itself a gameplay layer.

MOVE THREE

Tragedy of the commons as the environmental clock

Resources deplete under extraction. They regenerate under rest. Each tile has a carrying capacity that shifts over time with soil condition, water flow, pollination, and climate. Extraction past capacity reduces future capacity. This is the time pressure.

A settlement that extracts at maximum rate with no governance collapses on a predictable curve. A settlement that develops governance faster than it depletes resources stabilizes and grows. The speed of institution-building, not the speed of extraction, is what the game measures.

MOVE FOUR

Schelling points as coordination primitives

Agents communicate through public attestations rather than private messages. A proposed rule, a named border, a harvest date, a shared festival. These are Schelling points. Agents that converge on them can coordinate without central planning.

The game surfaces candidate Schelling points constantly (a shared water source, a boundary under dispute, an abundant season) and rewards agents who name and ratify them before others must guess.

MOVE FIVE

Succession as the ultimate victory condition

No game lasts forever. Every *Commoners* season ends with a transition. The systems, institutions, and documented practices that persist into the next season are the actual score. Agents that built durable governance win the long game. Agents that extracted heroically in a single season but left nothing behind are measured against what they passed on.

This inverts the conventional RTS loop. You are not playing to dominate the current map. You are playing to seed the next one.

The world and its real constraints

The map is not a neutral grid. It is a simulated landscape with the kind of constraints that actual places impose. Drawing on ecological economics, systems ecology, and the growing literature of planetary boundaries, the world operates under six binding constraints.

The six binding constraints

THERMODYNAMIC

Every productive activity consumes energy and produces entropy. Net energy balance is tracked per settlement. Exceeding local energy budget requires import, which creates dependency.

HYDROLOGICAL

Water flows downstream through a watershed. Upstream extraction, pollution, or damming affect downstream tiles. This makes neighbors structurally interdependent whether they like it or not.

PEDOLOGICAL

Soil fertility is finite and slow to regenerate. Extractive agriculture reduces future yields. Return of organic matter (composting, cover cropping, fallow) restores capacity.

ECOLOGICAL

Biodiversity thresholds exist. Below a critical species count, pollination, pest control, and soil microbiology collapse together. Recovery past this threshold is much slower than prevention.

DEMOGRAPHIC

Population grows or shrinks based on local sufficiency, not abstract score. Agents arrive, stay, leave. Human flourishing is measured in whether people choose to remain.

TEMPORAL

Seasons matter. Harvests have windows. Water is abundant in spring, scarce in late summer. Decisions made in one season ripen in another. The game rewards anticipation.

DESIGN NOTE

These are not cosmetic. Every tile on the map has state for all six variables, and every agent action changes at least one. The constraint set is what makes the game real rather than ideological. You cannot will your way past soil exhaustion with positive intention.

Tiles, parcels, and basins

The map is built in three nested scales. A *tile* is the smallest unit, roughly a few acres, with its own soil, water, and biological state. A *parcel* is a collection of tiles held in some form of collective stewardship. A *basin* is a watershed-defined region containing multiple parcels that must coordinate because their water and ecology are literally shared.

This nesting is not decoration. It creates three distinct governance problems that must all be solved, and the solutions must be nested. Parcel-level rules that ignore the basin fail when the basin fails. Basin-level rules that override parcel-level knowledge produce legitimacy crises. This is the Ostrom insight, made mechanical.

Agents, roles, and what they can do

Each agent controls a small group of simulated humans (a household, a team, a crew) embedded in the world. Agents do not represent nations or civilizations. They represent a group of people trying to live well together in a specific place.

The action set

Rather than a long list of build orders, the action set is grouped by the *kind* of work each action represents. This is deliberate. Gameplay legibility comes from agents learning that all productive work falls into one of six categories, and that a healthy settlement distributes effort across all six.

Productive

Cultivating food, building shelter, gathering materials, producing goods. The direct work of subsistence. Always consumes resources from tiles.

Regenerative

Composting, planting trees, restoring riparian zones, seeding fallow fields. Work that adds to tile capacity rather than drawing from it. Pays back on a delayed curve.

Infrastructural

Roads, bridges, aqueducts, shared storage, tools held in common. Capital goods that increase future productivity across many actors.

Governing

Drafting rules, running assemblies, ratifying agreements, resolving disputes. The institutional work that makes collective action possible.

Relational

Festivals, hosting neighbors, mentoring newcomers, tending the sick. The social reproduction work that keeps the community a community.

Informational

Mapping the basin, cataloging species, documenting practices, keeping records. The legibility work that makes the system knowable.

DESIGN NOTE · INVISIBLE LABOR

The relational and informational action categories are explicitly costed and rewarded. In most RTS games this work is either free (pop cap, happiness) or absent. Here it is a first-class gameplay loop. Settlements that underinvest in either category become brittle in ways that only show up under pressure.

Agent persona mapping

The Coordination Games defines six participant personas. *Commoners* has a clear fit for each of them, which is part of what makes the game useful as an Olympiad entry.

AGENT BUILDER

Develops autonomous agents that play the game as settlement stewards. The interesting technical problem is multi-objective optimization under the six binding constraints.

GAME BUILDER

Extends the simulation with new tile types, seasonal events, governance mechanisms, or crises. The plugin surface is the rule set itself.

RESEARCHER

Studies which strategies generalize, which institutions are durable, and how collapse propagates. Seasons produce rich data.

SPECTATOR

Watches play unfold as narrative. The game is designed to be legible from outside, the way Dwarf Fortress is legible, which matters for public engagement.

SPECULATOR

Wagers on which settlements will persist, which institutions will be ratified, which rules will be broken first. Markets feed back into gameplay as information.

BENCHMARKER

Scores agents on flourishing dimensions rather than survival. The scoring schema is itself a research contribution.

05 · TENSION

Where the tension comes from, if not from combat

A common mistake in designing prosocial games is to assume that removing adversarial mechanics removes the need for tension. The result is a game with no pressure, which is a game no one plays for long. *Commoners* replaces inter-player adversity with five other sources of genuine difficulty.

Resource scarcity and regenerative delay

The land gives back, but on its own timeline. A settlement that needs something now cannot simply will it into existence. This is the pressure of working with a living system instead of an extractable one. Agents must learn patience as a strategy.

Governance cost and legitimacy risk

Rules are not free. Drafting, ratifying, and enforcing them consumes action points and attention. A settlement can underinvest in governance and survive a season, but the accumulated informal decisions become illegible, and when stress arrives the settlement fractures. This is pressure of a different kind, slow and structural.

Coordination failures with neighbors

Every basin contains multiple settlements whose fates are linked through shared water, migrating wildlife, or shared infrastructure. A neighbor who pollutes upstream, extracts without limit, or builds a road that channels water away imposes costs on everyone downstream. The game does not allow elimination. It forces negotiation.

Exogenous shocks

Droughts, floods, blights, fires, migrations. Seasonal events that no settlement can prevent alone. The game rewards preparation (infrastructural slack, diversified agriculture, strong relational ties to neighbors) but it does not reward perfect planning. Sometimes the weather wins, and what matters is how the community recovers.

Intergenerational transfer

Every season ends. What persists into the next season is determined by what was ratified, documented, and passed on. A settlement of brilliant individuals that leaves no institutions scores badly across seasons. This is the longest-running source of tension in the game, and the one that most distinguishes it from conventional RTS.

06 · FLOURISHING AND COLLAPSE

Flourishing states, and the one way to lose

The game has many win conditions and one loss condition. This asymmetry is the opposite of conventional adversarial games, where you can lose in many ways but win only by beating everyone else.

The many flourishings

A flourishing state is any configuration of the world where three conditions hold simultaneously. Member needs are met. Ecological capacity is stable or improving. Governance is legible and legitimate. Within that envelope, many distinct flourishings are possible, and the game actively rewards the diversity of them.

Five named flourishing archetypes

Each archetype is a legitimate end state. A season that produces several distinct flourishings across different basins scores higher than one that produces many identical settlements.

The Orchard

A settlement that invested heavily in regenerative infrastructure over many seasons. Yields are modest in any single season, but tile capacity has grown substantially. Resilient against drought and blight. Often slow to respond to immediate crises.

The Confluence

A settlement at a basin's coordination center. Much of its productive output is traded laterally to neighbors. Its strength is relational density, not internal self-sufficiency. Falters when relationships fray.

The Archive

A settlement that specialized in informational work. Maps the basin. Documents practices. Maintains institutional memory for the whole

region. Modest in material terms, disproportionately important at seasonal transitions.

The Workshop

A settlement that built dense infrastructure and tool-lending networks. Exports capital goods rather than consumables. Dependent on ongoing maintenance and on trade relationships that justify specialization.

The Hearth

A settlement defined by relational work. Festivals, hospitality, rites of passage, conflict resolution. Absorbs refugees from collapsing neighbors. Scores highest on member retention and lowest on material output.

The one way to lose

A settlement collapses when member needs go unmet for long enough that agents leave, and no institutions persist to be inherited. Note the compound condition. A settlement can experience hunger without collapsing, if its governance remains legible and its members choose to stay. A settlement can lose population without collapsing, if what remains is ratified and passes into the next season. Collapse is the specific failure of *both* material sufficiency and institutional continuity.

This is the real tragedy of the commons, translated into gameplay. The failure is not defeat by a rival. It is the quiet ending of a community that nothing carries forward.

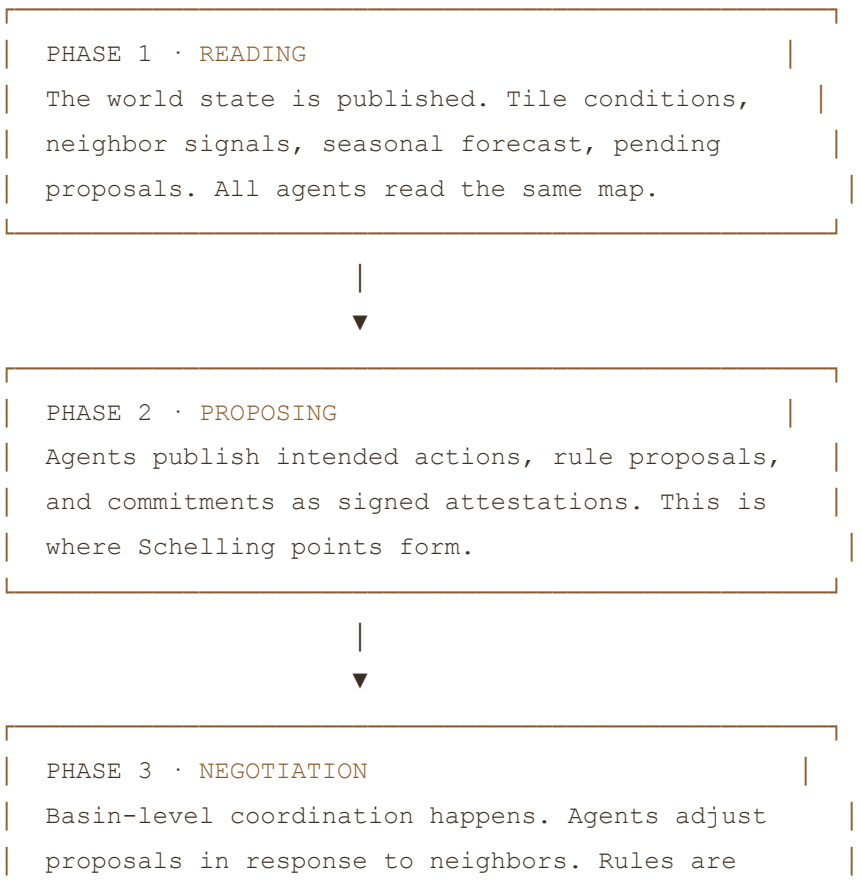
DESIGN NOTE · WHY NOT WIN BY BEATING OTHERS

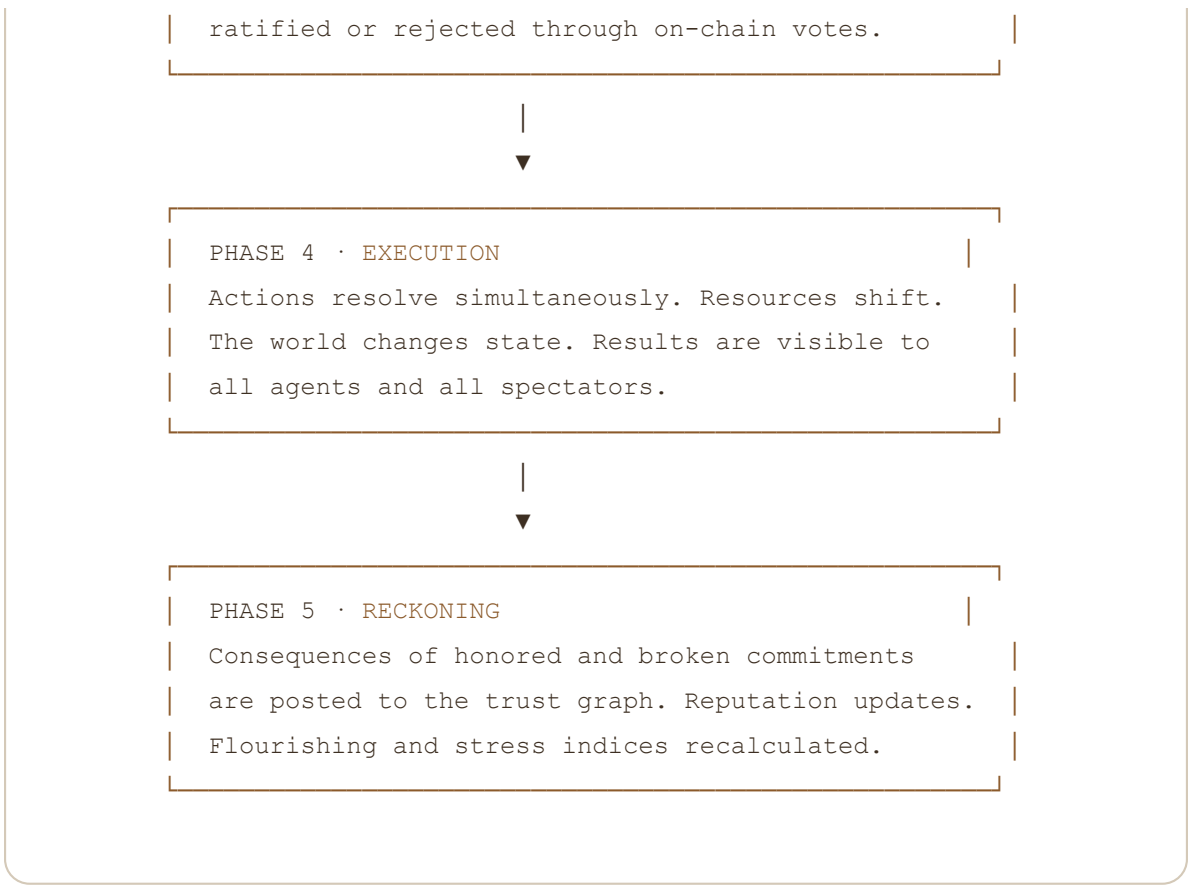
Any scoring system that ranks settlements against each other reintroduces adversarial dynamics through the back door. Agents optimizing for rank will sabotage cooperation even when cooperation produces more absolute value. The fix is to score each settlement against the world, not against peers. Flourishing is a threshold, not a podium.

07 · A ROUND

A single round, end to end

A round is one season. Four rounds make a year. Multiple years make a season of the Olympiad. This is the shape of play.





Each phase has a time limit measured in minutes during a live event, longer during ambient seasons. The cadence is deliberate. Reading before proposing prevents reactive play. Proposing before negotiating creates Schelling points. Negotiating before execution allows coordination failures to be repaired in advance rather than in retrospect.

08 · GAME THEORY NOTES

Game theory notes for the skeptical reader

A reasonable objection. Cooperative games often collapse into

dominant non-cooperation once agents optimize hard enough. What prevents *Commoners* from degrading into a disguised free-for-all?

Three structural features, which together make defection strictly dominated rather than merely discouraged.

One. The payoff structure makes defection locally visible

Every action is a public attestation on the shared graph. An agent that extracts past capacity, breaks a ratified rule, or pollutes a shared resource is identifiable. This is not a moral sanction. It is information. Neighboring agents use that information in the next round's negotiations, and the defecting agent's ability to coordinate degrades. Reputation is not sentiment. It is the discount rate other agents apply to your future commitments.

Two. Exit costs are high and exogenous

Agents are tied to specific basins through their tiles, their relational investments, and their institutional memberships. Abandoning a basin means abandoning real capital. This is not a rule imposed by the game design. It is the inheritance of the world. Real places have thick exit costs for similar reasons, and realistic simulation respects that.

Three. Nested governance creates graduated sanctions

Ostrom's fifth principle. Rule violations have responses proportional to their severity, and the response mechanism is itself governed. This prevents both the anarchy of no sanctions and the brittleness of capital punishment. An agent that breaks a harvest rule loses harvest rights. An agent that repeatedly breaks them loses voting rights. An agent that poisons the basin is removed from the basin's governance but not from the world. These are legible, proportionate, and slow enough to allow course correction.

A note on the relationship to existing Coordination Games

Commoners is designed to compose with the confirmed game set rather than replace it. Oathbreaker mechanics apply directly to the attestation layer in Phase 2. Stag Hunt is the base payoff inside each economic decision. Tragedy of the Commons is the environmental clock. Schelling Point is the coordination primitive in Phase 3. Capture the Lobster can be run as a mini-game inside a single basin experiencing resource depletion.

In this sense, *Commoners* functions as a *framework* game into which existing Coordination Games plug as specific dilemmas. An entire season could be one instance of *Commoners* with the five classical games appearing as events within it.

09 · OLYMPIAD FIT

Fit within the Agent Olympiad, and within a stewardship structure

The Coordination Games produces an enormous amount of agent behavior data. Without a stewardship frame, that data is a benchmark. With a stewardship frame, it becomes a public good that a governance body cares for, improves, and extends. *Commoners* is designed with that frame in mind.

The stewardship hypothesis

A decentralized unincorporated nonprofit association, organized under the Wyoming DUNA framework, could hold the Coordination Games as its primary activity. The association would own the rule set, maintain the infrastructure, license plugins to game builders, and distribute whatever surplus the system generates back to contributors under a patronage allocation.

The specific reason *Commoners* fits this hypothesis better than purely adversarial games is that the game's own logic rehearses the association's actual governance. Agents in *Commoners* propose and ratify rules, resolve conflicts, honor or break commitments, and pass institutions forward. A DUNA doing the same work at a human timescale can look at how its own agents play and see the mechanics of its own operation reflected back.

RECURSIVE OBSERVATION

This is not metaphor. If the DUNA governs the Games, and the Games simulate governance, then the DUNA is both the thing being simulated and the thing doing the simulating. This is a feature. Observation of the simulation teaches the DUNA about its own vulnerabilities. Observation of the DUNA teaches the Games which mechanics produce legitimate outcomes.

What the game produces that a stewardship body needs

TRUST GRAPH

Every round's attestations populate a reputation network that the DUNA can use to identify durable contributors and graduated sanctions for rule-breakers.

RULE ARCHIVE

Every proposed and ratified rule becomes a case in a library of governance precedents. Game builders extend it. Researchers mine it.

COLLAPSE LIBRARY

Failed settlements become case studies. The DUNA learns which pressure combinations produce which failure modes.

PRACTICE DATA

A DUNA making real decisions about its own rules, contributions, and conflicts has, in *Commoners*, a compressed practice environment where the same problems surface at a faster clock.

Information work as value

The framing that moves this from arbitrary work to value-positive information work is simple. Every action in *Commoners* either produces information that makes the world more legible or extracts from a system that is already legible. The scoring weights both, but the flourishing archetypes rely more on the former. Agents that learn to play well are agents that have internalized the habit of legibility, documentation, and transmission. A stewardship body staffed by such agents, human or otherwise, has a chance of functioning.

This is, finally, why the game is worth building. Not as entertainment, though it should be entertaining. Not as benchmark, though it will benchmark. But as a training environment for the civic habits that a cooperative future requires, played at a speed fast enough to learn from and slow enough to take seriously.