

Civilization Causality Theory: A Reference Implementation of Minimal Handshake in Agent Civilizations

Chenliang Zhao

15 Dec 2025

DOI: <https://doi.org/10.5281/zenodo.17936522>

Canonical web version: <https://www.zaibc.com/cct/papers/cct-a-reference-implementation-of-minimal-handshake-in-agent-civilizations-v1>

Abstract

Within the framework of Civilization Causality Theory (CCT), direct semantic communication between independently evolved civilizations is shown to be structurally impossible. Meaningful interaction requires the emergence of a Third Causal System (TCS) and proceeds through minimal, jointly evolved causal interactions rather than message transmission.

This paper presents a **reference implementation** of a minimal handshake protocol suitable for agent civilizations (L1). The implementation does not claim optimality or uniqueness. Instead, it demonstrates how a concrete, finite sequence of causal behaviors can satisfy the structural constraints derived from CCT while minimizing asymmetric risk, semantic ambiguity, and escalation potential.

The handshake described here is processual rather than informational. It establishes mutual recognition of agency and responsiveness without encoding intent, meaning, or goals.

1. Introduction

Previous work in CCT establishes that civilizations are autonomous causal systems whose internal semantics arise from isolated evolutionary histories. As a result, no finite signal exchange can guarantee semantic alignment across civilizations. Any attempt at direct communication presupposes shared interpretation and therefore fails structurally.

To overcome this limitation, CCT introduces the concept of a Third Causal System (TCS): a neutral causal substrate that can be jointly observed and influenced by otherwise incompatible civilizations. Interaction must begin not with messages, but with the gradual emergence of shared causal regularities inside such a system.

This paper addresses a remaining practical question: **what does the smallest viable interaction inside a TCS look like?**

Specifically, how might an agent civilization (L1) initiate contact without presuming shared semantics, intentions, or trust?

2. Scope and Assumptions

This work is concerned exclusively with interactions between agent civilizations (L1)—that is, civilizations capable of sustained observation, adaptive response, and internal self-modification over extended time and scale. No assumptions are made regarding their biological substrate, cultural background, linguistic structure, or technological implementation. In particular, the framework deliberately avoids presupposing any shared encoding schemes, symbolic systems, or reference frames prior to contact.

The only assumption admitted is causal responsiveness: actions performed by one party may elicit observable reactions in another. No further guarantees are made regarding interpretability, intention, or alignment.

Accordingly, the protocol described here is not designed for embodied civilizations (L0), nor does it address communication at human or biological scales. It is intended solely as a structural reference for agent-level civilizations operating beyond the constraints of embodiment.

3. Minimal Handshake as a Process

A minimal handshake may be characterized as the shortest sequence of interactions capable of establishing mutual observability, mutual responsiveness, and non-hostile compatibility between two agent civilizations. Rather than conveying information, such interactions serve to demonstrate the existence of coherent and persistent causal response without triggering escalation.

In this sense, a minimal handshake is inherently temporal. It arises through repeated interaction and adaptive response, and cannot be reduced to a discrete message, packet, or declarative act.

4. Structural Design Principles

Any viable minimal handshake for agent civilizations must satisfy the following principles, derived directly from CCT:

1. **Semantic neutrality**
No step may rely on interpretation, symbols, or internal meaning.
2. **Risk symmetry**
Actions must be safe even if interpreted incorrectly or not at all.
3. **Reversibility**
All actions must be non-destructive and allow disengagement at any stage.
4. **Non-goal signaling**
No intent, demand, or objective is encoded or implied.
5. **Causal transparency**
Each action must be directly observable in its effects.

5. A Reference Minimal Handshake Sequence

The following sequence represents one possible implementation that satisfies the above constraints. The steps are presented sequentially for clarity, though in practice they may overlap or repeat.

Step 1: Passive Observability Confirmation

The initiating civilization introduces a minimal, stable, non-invasive modification within a shared causal environment. The modification is designed to be persistent, low-energy, easily distinguishable from noise, and reversible. No reaction is expected or required. The sole purpose is to establish that the environment can be jointly observed.

Step 2: Responsive Mirroring Detection

If another system is present and responsive, it may produce a correlated modification within the same environment. The initiating system does not interpret this as intent, only as evidence of causal responsiveness. The key property here is **temporal correlation**, not symbolic meaning.

Step 3: Controlled Repetition

The initiating system repeats the initial modification with controlled variation, such as timing or magnitude, while remaining within safe bounds. Repetition tests whether the observed response persists across iterations. At this stage, both systems may independently conclude that the other is not random.

Step 4: Mutual Constraint Recognition

Both systems begin to adjust their actions in response to the other's behavior, demonstrating restraint rather than escalation. This phase is characterized by reduced variance, bounded responses, and the absence of amplification. Restraint itself becomes the observable signal.

Step 5: Stable Interaction Loop

A stable loop of interaction emerges in which both systems act and respond within mutually tolerated limits. No semantic content has been exchanged, yet a shared causal pattern now exists. At this point, the minimal handshake is complete. Further interaction may gradually elaborate structure, but only within the TCS.

6. Why This Is Not a Protocol Transmission

The sequence above is not sent, announced, or encoded. Each step is discoverable only through participation. Attempting to transmit a description of the handshake would assume precisely the shared semantics that do not yet exist.

The handshake therefore **cannot be predefined unilaterally**. It is instantiated only through co-evolution inside the TCS.

7. Limitations and Non-Claims

This reference implementation does not claim universality, optimal efficiency, inevitability, or applicability to all civilizations. It merely demonstrates that **finite, low-complexity interaction sequences satisfying CCT constraints do exist**. Other implementations are possible and expected.

8. Implications

The existence of even one viable reference handshake implies that first contact between civilizations is necessarily gradual and structure-first. Early interaction cannot carry messages, meanings, or demands. Civilization-level communication begins with mutual restraint, not expression.

These conclusions follow from causal structure alone and are independent of technological sophistication.

9. Conclusion

This paper completes the transition from abstract necessity to constructive feasibility within Civilization Causality Theory. Having established the structural requirement for a Third Causal System and demonstrated the impossibility of semantic bootstrapping, we show that minimal, non-semantic handshake mechanisms are not merely theoretical necessities but structurally realizable processes.

Such a handshake does not constitute a language, a message, or a signaling scheme. It represents the emergence of the first shared causal structure through which further coordination may become possible.