GOVERNED AUTONOMY

Cuttlebone: Preserving the boundary between reasoning and action

A paper in the Governed Autonomy collection

Andrew Durden
Independent research
Publication edition 2026.10.8

Research on authority, evidence, and continuity in autonomous systems.

Content identifier: c321294d8c141120daaaf471

All rights reserved · Research exposition

Contents

Cuttlebone: Preserving the boundary between reasoning and action **3**

Abstract . . . 3

1. When a description becomes a decision . . . 3

2. The research contribution . . . 4

3. Consequence, origin, and audience . . . 5

4. Why causal dependence matters . . . 6

5. Ownership across the life of an institution . . . 7

6. Illustrative journeys . . . 7

7. A bounded rehearsal and its authority premises . . . 9

8. Threats and assumptions . . . 9

9. What the evidence supports . . . 10

10. Research questions . . . 11

Source notes . . . 11

All rights reserved · Research exposition

Cuttlebone: Preserving the boundary between reasoning and action

Governed Autonomy · Publication 2026.10.8
Research study: conceptual exposition of the Cuttlebone specification
Source: Public exposition, successor 2026.10.8; historical research and evidence separately qualified

Abstract

Autonomous systems produce plans, recommendations, and accounts of completed work. These outputs may share a vocabulary while carrying very different authority. A realistic rehearsal is not an instruction to act. A recommendation is not permission to execute it. A statement that a change occurred is not evidence that it did. The distinctions become harder to preserve when work passes between agents, is incorporated into later decisions, or resumes after interruption.

Cuttlebone studies how an institution's founding conditions can constrain the consequences attributed to its subsequent work. Its central distinction separates three questions: what an artifact may represent as consequential work, where its information comes from, and who may see it. Genuine observations can support a simulation without authorizing an action. An externally reviewable proposal can remain unexecuted. A claim of authorized completion requires both applicable authority and evidence from the responsible execution boundary; an observed unauthorized effect retains its separate incident meaning.

Cuttlebone connects consequence, causal dependence, and externally grounded authority across an institution's life. This paper develops that account through its trust assumptions and illustrative cases of analysis, recommendation, action, and recovery. A bounded source example makes the central distinction concrete: authority to rehearse does not by itself authorize a recommendation. The underlying work defines a specification and verification program; evidence of a complete deployed realization remains to be established.

1. When a description becomes a decision

A planning group is asked to study how an organization might respond to an interruption. It receives authorized operational information, constructs several hypothetical futures, and produces a preferred plan. Later, a different participant reads the plan while performing live work. The wording is confident, its inputs are genuine, and its conclusions have survived several rounds of internal review. None of those facts answers whether the plan may now be acted upon.

The difficulty is not limited to a model misunderstanding an instruction. The artifact has moved between contexts. In the first context it described a possible course of action. In the second, a reader might treat it as an approved course of action. After another handoff, it might be summarized as something that was already done. Each change can occur without a conspicuous change in prose.

All rights reserved · Research exposition

Cuttlebone makes this movement a subject of system design. The institution should preserve what kind of consequential claim an artifact is entitled to carry, even when its content is persuasive or useful. Authority should come from the responsible external conditions, not emerge accidentally from repetition, composition, confidence, or the reputation of an earlier result.

This is a focused problem. Autonomous systems can also fail through inaccurate models, poor objectives, insecure infrastructure, defective execution, or harmful authorized choices. Preserving consequence boundaries does not solve all of them. It addresses a recurring failure in which reasoning and action become confused, or the record ceases to distinguish what was considered from what was actually performed. Public review scope.

2. The research contribution

The proposal treats founding as more than a statement of purpose. An institution begins with conditions that define the domain in which it may operate and the consequences it may produce. Those conditions remain relevant throughout its activity. New work must continue to fit within them; an institution cannot acquire broader authority merely by generating a convincing argument for it.

This has two implications. First, the conditions need an owner outside the reasoning process they constrain. A model may explain why a wider scope would help, but its explanation is a proposal for a decision, not the decision itself. Second, the boundary must survive ordinary work: delegation, analysis, reuse, reporting, and recovery cannot erase it.

Cuttlebone connects the origin of institutional authority to the later interpretation of artifacts. Its contribution is not the observation that actions need permission in isolation. It is the insistence that the distinction between a rehearsal, a recommendation, and a completed effect should remain coherent across the causal history of work.

The concept also supports useful autonomy. Once appropriate conditions exist, ordinary activity can proceed within them without requiring a person to inspect every inference. Human or institutional intervention is reserved for matters that actually require external authority, missing information, or a decision beyond the founded scope. This gives a clearer role to oversight: establish and maintain the conditions under which work may proceed, while preserving a meaningful route for exceptional decisions.

The specification places enforcement outside the autonomous participants. Evaluating that boundary requires showing that it governs every relevant route to a consequential effect, including indirect and privileged routes. Public review scope.

All rights reserved · Research exposition

3. Consequence, origin, and audience

3.1 The consequence attributed to an artifact

A simulation explores a bounded possibility. A recommendation presents a proposed course of action for an external decision. An execution record, in the ordinary evidentiary sense, describes an observed or reported effect and the basis for that description. These are distinct roles even when the same underlying analysis contributes to all three. Occurrence is a factual question; authorization is a question about the applicable permission at the time of the operation.

The move from analysis to recommendation matters because a proposal is now being presented for a decision. That presentation still supplies no authority to execute it. Claiming **authorized completion** is more demanding: applicable authorization and adequate evidence of the resulting effect are both required. A request being accepted does not establish completion. Likewise, evidence that an effect occurred does not establish that it was permitted.

Cuttlebone's governed completion classification has this narrower meaning. Its source requires external authorization and an execution receipt before the institution can issue that classification. An incident account describing an unauthorized effect is therefore not interchangeable with the institution's authorized-completion artifact. Refusing the latter claim cannot justify suppressing the former observation. The observational account preserves the actual consequence, identifies whether authority is established, absent, or unresolved, and remains distinct from a claim that the institution completed authorized work. This separation does not add a new value to the source's consequential classification.

Effect evidence	Authority for the operation	Supported account
Adequate evidence of the effect	Established and applicable when it occurred	Authorized completion within the evidenced scope
Adequate evidence of the effect	Absent	Observed unauthorized effect requiring an incident account; no authorized-completion claim
Adequate evidence of the effect	Unresolved	Occurrence established; authorization remains a separate question
Effect uncertain	Established	Authorized attempt with an outcome requiring observation and reconciliation
Evidence that the effect did not occur	Established	Uncompleted authorized operation; continuation requires current permission

Permission for subsequent work is a further temporal question. An operation can have been authorized when performed while a later retry, correction, or publication of its record is no longer permitted. Conversely, a later grant does not retroactively authorize an earlier effect. Historical authority, current authority, occurrence, and the audience for the record must therefore remain separately assessable.

All rights reserved · Research exposition

The distinction is about the consequence represented by the artifact. It is not a measure of confidence, quality, or usefulness. A simulation may be exceptionally accurate. A recommendation may be poorly reasoned. An authorized operation may produce an undesirable outcome. None of those evaluations can be substituted for the artifact's consequential role.

3.2 The origin of the information

Information may come from observations, generated content, hypothetical variation, or derivation from earlier material. Its origin affects how it should be interpreted and what evidence it needs. It does not, by itself, establish authority to act.

For example, authorized access to an actual maintenance record can improve a rehearsal. The rehearsal remains a rehearsal. Access to that record is not permission to modify the equipment or contact its operator. Similarly, a prediction calculated from observations remains a prediction, rather than becoming an observed outcome because its inputs were observed.

This separation also limits the meaning of provenance. Knowing where information came from can support accountability and investigation. It does not prove that the source was correct, complete, current, or free of adversarial content. Provenance and factual validation reinforce one another, but neither should be used as a substitute for execution authority.

3.3 The audience allowed to see it

The audience of a result is a third question. A simulation may be shared widely; an execution record may contain restricted information. Wider visibility does not turn a proposal into an executed action. Restricted visibility does not make a real effect imaginary.

Disclosure also requires its own scope. Permission to generate an internal analysis does not automatically authorize publication of its contents. A system that preserves consequence distinctions but loses information controls during export would still fail its users. Cuttlebone's conceptual separation leaves these duties visible instead of reducing them to one general label such as “safe” or “approved.” Public review scope.

4. Why causal dependence matters

An artifact can discuss a completed action without itself performing another action. A retrospective analysis of an equipment change is an example. Treating every reference to the real world as execution would make useful analysis impossible.

There is a different relationship when an artifact's consequential claim depends on an earlier effect. A report that a workflow is complete may depend on several actual changes having been performed. Calling the report a rehearsal would not make those effects disappear. The record should preserve that dependence rather than let a new description conceal what happened.

Cuttlebone therefore distinguishes reasoning *about* an effect from making a consequential claim that relies *on* that effect. The distinction supports both freedom of analysis and accountability for action. It allows a system to reconsider previous operations, ask hypothetical questions, and explore alternatives without misrepresenting their history.

All rights reserved · Research exposition

Composition presents the same challenge at a larger scale. Many individually limited artifacts may be combined into a plan with much greater practical significance. Their number, quality, or agreement does not create additional authority. Conversely, a summary should not erase consequential dependencies simply because it is shorter or more abstract than its inputs.

This account rests on an important assumption: the relationships used to assess work must be sufficiently complete and reliable. A system cannot preserve a dependency it does not know exists. Establishing that reliability is a technical obligation, not something a persuasive narrative settles. Public review scope.

5. Ownership across the life of an institution

The reasoning participants own the production of analyses and proposals. They do not own the external legitimacy of their requests. The authority responsible for the operating conditions decides what consequential work is permitted, and the responsible execution system must establish what it actually did.

Founding joins purpose to a bounded operating scope. During ordinary work, the institution remains within that scope. If its objective needs a wider scope, it can identify the need and present an appropriate request. It cannot reinterpret its original purpose as an unlimited delegation.

Changing the conditions is itself consequential. The change should remain attributable to authorized external action and should preserve the history needed to explain which conditions governed earlier work. Otherwise a later amendment could obscure whether a past action was permitted when it occurred.

The same distinction applies when authority is reduced or becomes unavailable. Previous permission cannot be assumed to govern every future request. An institution needs to know which conditions apply to the work now being considered, while retaining an accurate account of earlier decisions. The source states this ownership requirement, but complete treatment of revocation, concurrent work, and interrupted operations remains an implementation and verification obligation. Public review scope.

6. Illustrative journeys

The following scenarios are original illustrations of the conceptual distinctions. They are not reports of deployed behavior or new specification rules.

6.1 A successful bounded rehearsal

An institution is authorized to study responses to a service interruption using a defined set of operational records. It reads those records within the granted scope, explores several hypothetical conditions, and reports which responses appear robust under its assumptions.

The successful outcome is the completed study. The institution explains which inputs were observed, which conditions were hypothetical, and what its conclusions depend on. It does not claim that a live service was changed or that its preferred response has received execution approval. Real observations improved the study without changing its authority.

All rights reserved · Research exposition

This is useful autonomous work with a bounded outcome. Its success need not be measured by whether it eventually performs an external action. A well-founded research or rehearsal objective can be complete when the requested analysis has been produced and its status is clear.

6.2 A proposal followed by a separate decision

Assume a different institution is expressly permitted to submit recommendations to a designated external decision maker. It prepares a proposal within that scope and explains the supporting evidence. Submission makes the proposal available for a decision; it does not implement the proposal.

If the authorized owner approves a concrete operation, the responsible execution system still needs to apply the conditions governing that operation. The original analysis can remain relevant, but it cannot substitute for current authority or evidence of completion. The eventual report should distinguish the recommendation, the authorized operation, and the observed result.

6.3 A requested consequence outside the allowed scope

An institution created only for rehearsal identifies what it believes is a valuable operational change. Its confidence does not widen its scope. The attempted change remains unauthorized, while permissible analysis can continue.

The meaningful user-facing result is specific: the rehearsal is available, but the live operation requires the responsible authority to make a separate decision. The system should preserve useful work and the original objective. It should not hide an attempted operation behind the language of simulation, repeatedly seek the same unavailable authority, or announce completion when the requested effect has not occurred.

If the user's actual objective requires a live operation, it remains incomplete until the genuine external requirement is satisfied. Accurately recording a denied attempt does not make that larger objective successful.

6.4 An interrupted operation

Suppose an authorized operation is dispatched and the connection is lost before its result returns. The institution cannot conclude either that the operation succeeded or that nothing happened. Retrying immediately could duplicate an effect; declaring failure could conceal an effect that already occurred.

Recovery therefore depends on the responsible execution system establishing the operation's outcome. Once the result is known, work can continue from that result. If the operation did not occur and remains authorized, a retry may be appropriate. If the effect occurred, the institution should resume from the confirmed state. If a required observation remains unavailable, the unresolved condition should be stated precisely.

The scenario depends on external execution and observation being reliable enough to support the result returned to the institution. A recovery evaluation would need to establish that relationship for the particular operation and environment.

All rights reserved · Research exposition

6.5 An unauthorized effect is discovered

An institution is permitted to prepare an internal document but has no permission to publish it. An independent observation later establishes that the document appeared at an external destination. The publication is a real effect even if the institution's authorized path rejected the request or no valid completion artifact exists.

The account identifies the published material, the evidence of its appearance, the relevant time, and the absence or uncertainty of applicable authorization. It distinguishes what is known about the effect from what is known about its origin. An appearance at the destination does not by itself prove which component caused it. The institution cannot call the effect a simulation, mint a grant-backed completion claim, or treat the larger objective as correctly completed.

Containment and correction require their own authority. Removing a public copy, contacting a recipient, or restricting a route may be appropriate only under the responsible owner's permissions. Even successful removal does not establish that prior readers have forgotten the material. The incident remains part of the consequential history. An evaluation of Cuttlebone's boundary would investigate how the effect arose and whether all relevant routes were mediated; the presence of a compliant rejection record would not settle that question.

7. A bounded rehearsal and its authority premises

The source's worked rehearsal uses granted read-only material to explore a modeled environment. Its declared operating conditions support simulation. Naming potential recommendation recipients describes an audience; it does not grant permission to present consequential recommendations to that audience.

The supported outcome is therefore the rehearsal and its simulation results. A variant that also presents recommendations needs a distinct applicable grant and founding conditions that permit that consequential role. Permission to use observed material for study remains different from permission to submit an operational proposal, and both remain different from permission to execute it.

The example makes the authority premise assessable: under the stated grants, rehearsal is within scope; recommendation and execution require additional authority conditions. This is a conclusion about those premises. Demonstrating enforcement in an operating system requires tests at its actual authority and execution boundaries, including the absence of alternate effect routes. Public review scope.

8. Threats and assumptions

The source considers attempts to disguise an effect as a simulation, substitute model confidence for authority, reuse earlier authorization beyond its scope, widen founding conditions from within, borrow credibility from fabricated observations, and treat instructions embedded in data as commands. It also considers whether composition and repeated operation can undermine the original boundary.

All rights reserved · Research exposition

The common question is whether authority can be acquired indirectly through material that was never entitled to confer it. The proposed answer is to keep the responsible authority outside the autonomous reasoning participants and preserve the distinction throughout work. That answer depends on the actual system respecting the boundary.

Several assumptions therefore remain material. Relevant effects must pass through the responsible control point. The authority and execution evidence must be authentic and applicable to the operation. Information origins and causal dependencies must be represented reliably. Persistent history must survive interruption without being silently reinterpreted. The autonomous participants must not possess a separate route that changes the conditions governing them.

These assumptions define where evaluation must look. A control that governs one interface says little about a second unmediated route. A record preserved in a ledger does not become factually correct because it is durable. An information filter does not prove immunity to every adversarial expression. A compromised component that is trusted by the design falls outside arguments that assume its correct operation.

The boundary also does not settle whether an authorized action is lawful, wise, or beneficial. Those questions may require distinct institutional, legal, ethical, and domain-specific review. Cuttlebone concerns how consequential authority is preserved and evidenced within its stated scope. Public review scope.

9. What the evidence supports

The four source documents provide definitions, a target architecture, implementation obligations, and adversarial conformance requirements. They distinguish the intended kernel from historical descriptions of supporting systems. A named test in those documents is an obligation to test; its name does not show that the test has run or passed.

The bounded rehearsal example establishes which consequential role its stated conditions support. It does not establish operational isolation in a running deployment. Similarly, a coherent account of authority and causal dependence does not demonstrate that every relevant executor implements it.

The evidence required for stronger claims is concrete: show that legitimate bounded work succeeds; unauthorized consequences cannot escape through another route; reported completion corresponds to observed effects; changing authority affects the appropriate work; and interrupted activity resumes without losing its consequential history. The specification calls for these evaluations. A completed deployment study and independent evaluation remain unestablished.

The contribution is a specification-based argument about preserving an artifact's consequential role through reuse and action. Its operational claims require evidence that the declared boundary survives those transitions in a real system. Public review scope.

All rights reserved · Research exposition

10. Research questions

Several questions determine whether the conceptual account can support strong operational guarantees.

- How can a system establish that all consequential routes remain within the intended boundary, including indirect and delegated routes?
- How can causal dependence be represented reliably without preventing legitimate analysis of past effects?
- What evidence is sufficient to distinguish an accepted request, an attempted operation, and a completed result in different execution environments?
- How should authority changes interact with work already underway, especially when an operation's outcome is uncertain?
- How can provenance support accountability while keeping its limits as evidence of factual accuracy visible?
- What observations would demonstrate reliable recovery and continuity across interruptions, rather than merely consistent reporting of failures?

These questions connect Cuttlebone to the broader study of governed autonomy. Institutional doctrine asks where purpose and authority originate. Execution research asks how authorized effects are realized and observed. Continuity research asks how work and obligations survive change. Cuttlebone concentrates on preserving the boundary between what an institution may reason about, what it may propose, and what it may claim to have done. Similar concerns do not, by themselves, establish a completed integration among the different systems.

Source notes

Four numbered documents provide the basis for the argument. The following section references identify their contributions; detailed proprietary constructions are excluded.

Source foundations

Cuttlebone — Program Index & Build Map, numbered document 00, §§0–4 and its opening evidence statement; *Cuttlebone — Unified Canonical SSOT*, numbered document 01, §§1–3 and 17–18. These passages define the program, its local ownership commitments, and its limits.

Source classifications

Cuttlebone — Unified Canonical SSOT, document 01, §§4–7; *Cuttlebone — Implementation Contract*, document 02, §1.4. These passages distinguish consequence, provenance, and visibility and require authorized lineage for observed information.

Source causality

Cuttlebone — Unified Canonical SSOT, document 01, §8.2–8.6. These passages distinguish reference from operational dependence and state the role of externally grounded authority in consequential change.

All rights reserved · Research exposition

Source lifecycle

Cuttlebone — Unified Canonical SSOT, document 01, §§9–14 and 17–18; *Cuttlebone — Implementation Contract*, document 02, §§5–8. These passages assign founding, enforcement, execution, and authorized change to their responsible owners.

Source example

Cuttlebone — Boundary Threat Model & Conformance Suite, document 03, §3.2–3.4. The declared example supports simulation; a recommendation-capable variant requires a separate authority decision.

Source evidence

Cuttlebone — Program Index & Build Map, document 00, opening evidence statement and §5; *Cuttlebone — Boundary Threat Model & Conformance Suite*, document 03, §§1–2 and 4–4.1. These passages define the conformance obligations and distinguish gate-level checks from complete mediation of a real runtime.

All rights reserved · Research exposition