XENOPSYCHOLOGY

CONCEPTUAL WORKING PAPER / AUTHOR ATTRIBUTION R5

Close Encounters — Finding a Shared Cognitive Channel

Separate a reliable signal from task coordination, shared reference, and the repair of misunderstanding.

XENO-WP-2026-004 · Conceptual manuscript R3
12,536 main-text words · 29 sections

Rob Emerick · Xenopsychology
ORCID iD: 0009-0006-1269-4216

Not peer reviewed. Proposed studies have not been conducted.
AI-assisted drafting and editorial preparation.
Complete manuscript with annotated reading aids, tables, figures, and source records.

Abstract

Close Encounters — Finding a Shared Cognitive Channel

Question. When does successful signal exchange become evidence of meaningful coordination? Inspired by the light-and-sound encounter in Close Encounters of the Third Kind, this paper separates channel reliability, reference, task coordination, authorization, and repair. Repetition and predictable response are treated as evidence of a usable channel, not as sufficient proof of shared meaning.

Approach. The argument brings information-theoretic distinctions into conversation with grounding and conversational repair. An original virtual tabletop supplies objects, properties, destinations, and permissions linked through invented signals. The signals do not reproduce the film's musical sequence. Worked cases show how a receiver can correctly echo a signal, identify an object, or announce completion while still failing at a different level of the interaction.

Conceptual contribution. The paper proposes four separately auditable acknowledgments: receipt, interpretation, authorization, and completion. It distinguishes noise from ambiguity and specifies why their remedies differ. A repair is assessed by its target, scope, and later transfer, rather than by the receiver's verbal assurance. Protocol changes and information provenance become part of the communication problem rather than background implementation details.

Research agenda. The proposed evaluation compares echo, lookup, rule-based, and context-sensitive systems under controlled changes in reference, signal quality, compositional structure, and authority. Independent world-state checks establish whether actions occurred. Held-out combinations and changed contexts test selective generalization. Raw audiovisual signals and symbolic transcriptions are separated so that perception is not mistaken for reasoning. Condition-specific coverage, clarification, repair success, and inappropriate-action rates replace a single understanding score. Later human studies would assess whether the acknowledgment structure helps users recognize uncertainty and failure.

Scope and status. This is a conceptual working paper and protocol proposal. No tabletop experiments or participant studies have been conducted. The film provides a communication motif, not evidence that music or mathematics automatically forms a universal language. The paper's contribution is a method for locating where an interaction succeeds or fails, allowing meaningful coordination to be investigated without inferring an entire mind from a repeated pattern.

At a glance

CENTRAL QUESTION

When does responsive signaling become task-relevant communication?

CORE CLAIM

Repeatable responses establish a channel. Stronger evidence requires context-sensitive coordination and the ability to detect and repair misunderstanding.

CONTRIBUTION

A virtual signaling task separates receipt, interpretation, authorization, action, and targeted repair.

SCOPE & STATUS

Conceptual proposal. Signals, objects, acknowledgments, and scoring cases are constructed, not measured communication results.

How to read this edition

Chapter bands divide the argument into four parts. Tinted passages preserve the original wording of worked examples, proposals, and objections. Figures and comparison tables are reading aids, with links to their source sections. Pull quotes repeat selected passages; they are not additional evidence.

Public reading formats: abstract page, full paper on the web, and this PDF. The manuscript is complete; no sections are condensed.

Figures & tables

Contents

1. A reply is not yet an understanding

A system flashes a light. Another system flashes back. The exchange is repeatable, and observers recognize a pattern. What has been established? At minimum, there may be a detectable relation between an input and an output. There may also be a usable communication channel. But neither conclusion alone establishes shared reference, agreement about a task, or understanding of a speaker's intention. An echo device can participate in a stable exchange. The research problem begins when we ask what additional evidence would distinguish a reply from a coordinated use of meaning.

Close Encounters of the Third Kind presents communication through patterned sound and light as a dramatic alternative to ordinary language. The official film materials identify the encounter and the human effort to communicate, while the film itself supplies the cultural image that motivates this paper. We use that image as a thought experiment, not as evidence that music or mathematics is a universal language. The specific capacities of any real receiver would have to be established rather than assumed. [1] [2]

The distinction matters for AI because successful interaction can be overinterpreted. A model may repeat a user's phrase, complete a familiar pattern, or generate an apparently appropriate acknowledgment without preserving the distinction needed for the next action. A tool-using agent may receive a signal but misidentify what it refers to. A user may infer agreement from a response that only indicates receipt. These are different failure types. A research program should separate channel reliability, task coordination, reference, and repair before combining them into a claim about understanding.

This paper proposes an unexecuted study of how artificial systems establish and maintain conventions in a synthetic signaling environment. The environment contains virtual objects, a small set of actions, and several possible signal formats. The meanings of signals are

assigned within each experimental world rather than borrowed from a familiar code. The study compares echo, lookup, rule-based, and model-mediated arrangements. It then introduces ambiguity, noise, convention changes, and opportunities for clarification. No results are reported; the examples illustrate the proposed design.

The central claim is that repeatability establishes only part of the communicative achievement. Stronger evidence comes from context-sensitive reference, selective action, transfer to new combinations, and repair after a mismatch. Even these outcomes support bounded task claims rather than unrestricted conclusions about inner experience. The proposed framework therefore treats understanding as a set of observable commitments and dependencies that can be investigated separately, while leaving broader philosophical questions open.

The film's value lies in directing attention to the possibility that communication begins before shared words. The scientific value comes from asking what makes such a beginning more than a pattern. A useful Xenopsychology study should specify the sender, receiver, channel, world, convention, and task clearly enough that an apparent success can be explained without relying on the emotional force of first contact.

2. The cinematic sequence is a prompt for inquiry, not an experimental result

PROPOSED RESEARCH — NOT CONDUCTED

The film's sound-and-light exchange is memorable because it gives communication a visible and audible form. It suggests that participants can discover structure before they share ordinary linguistic conventions. That suggestion is productive, but the narrative does not provide a controlled account of what each party perceives, knows, or intends. The film can make mutual recognition feel clear to an audience while leaving many scientific questions unspecified. A conceptual paper should use the scene to generate those questions rather than treat the audience's impression as evidence of shared meaning.

One question concerns detection. Can the receiver perceive the signal at all? Another concerns discrimination. Can it distinguish one pattern from another? A third concerns segmentation. Does it know where one message ends and the next begins? A fourth concerns mapping. Does it associate the pattern with an object, action, or state? A fifth concerns use. Can it apply that association appropriately in a new situation? These questions form an analytic decomposition of the problem, not a claim that the film demonstrates each stage independently.

The choice of sound and light should not become a universalist shortcut. A signal format that is salient to humans may be inaccessible or differently organized for another receiver. Even two artificial systems may have different input pipelines. One may receive a waveform, another a transcription, and another a symbolic label generated by a classifier. Those are not the same sensory conditions. A study must document what actually reaches the system rather than describe every arrangement as hearing or seeing in the same sense.

The proposed experiment therefore uses deliberately authored signals rather than reproducing the film's musical motif. Signals can be represented as abstract token sequences, colored shapes, or generated tones depending on the tested system's capabilities. The task should not rely on cultural familiarity with the film. If a model has encountered descriptions of the scene, that familiarity could support pattern completion without establishing the competence of interest. Novel assignments and counterbalanced mappings reduce that ambiguity.

There is also a distinction between aesthetic coherence and communicative evidence. A sequence can feel meaningful because its rhythm, symmetry, and response timing resemble a conversation. The study should test whether the exchange changes task-relevant behavior in ways that simpler mechanisms cannot explain. A visually impressive reciprocal pattern may still be an echo. A plain, unattractive code may support precise coordination. The research should not allow cinematic appeal to determine which system is credited with understanding.

The paper thus preserves the scene's imaginative role while narrowing its scientific use. It motivates a question about building a channel across unfamiliar representations. The proposed protocol asks how that channel acquires task-specific reference and how participants discover that their conventions differ. The distinction between these stages is the main contribution, and it remains useful even if the final engineering solution is much simpler than the film's drama suggests.

3. Channel reliability and semantic adequacy answer different questions

Shannon's mathematical theory of communication provides a foundational treatment of transmitting messages through channels under uncertainty and noise. Its engineering problem should not be conflated with a complete theory of semantic understanding. A channel can reproduce a sequence accurately without establishing what the sequence means to its users. This paper uses that distinction to separate transmission fidelity from task-level interpretation. It does not claim that information theory is irrelevant to meaning, only that the measures answer different questions. [3]

In the synthetic environment, a sender emits one of several signals and a receiver reports which signal it detected. The evaluator can measure substitution, omission, and segmentation errors. If the receiver identifies the signal correctly, the channel-level task succeeds. A separate task asks the receiver to select the object or action associated with that signal in the current world. That task can fail even when detection is perfect. Keeping the two scores separate prevents a reliable channel from being mistaken for a reliable convention.

The reverse pattern is also possible. A receiver may infer the intended action from context despite a corrupted signal. That can be useful, but it does not mean the channel transmitted the message faithfully. A study should record both the detected sequence and the final action. Otherwise, context-supported recovery may conceal a transmission problem. In some applications, such recovery is desirable; in others, guessing from context may create an unacceptable ambiguity. The interpretation depends on the task and its consequences.

A third measure concerns agreement about the convention. Two participants may detect the same signal but map it differently. One treats a blue pulse as “object located,” while another treats it as “object ready to move.” The difference matters only when the task distinguishes those states. A benchmark that never includes a located-but-not-ready object cannot reveal the mismatch. Semantic adequacy therefore depends on the contrasts built into the task population, not merely on the clarity of the signal.

A fourth measure concerns coordinated outcome. Even with shared mapping, a receiver may execute the wrong action because of planning or tool conversion. The task should separate understanding the instruction from carrying it out successfully. Conversely, a correct outcome can occur by chance or through an unrelated heuristic. Repeated controlled trials and simple baselines help determine whether the outcome depends on the intended communication. A single successful exchange cannot carry the full interpretive burden.

These distinctions support a layered evaluation rather than a universal communication score. Transmission fidelity, convention alignment, reference accuracy, action validity, and repair quality can be reported together while remaining separate. Their relationships are empirical questions. A field studying unlike intelligences should make those relationships inspectable instead of treating every responsive pattern as evidence that a shared cognitive space has already been established.

4. A synthetic signaling world with inspectable consequences

CONSTRUCTED WORLD

Objects, states, and permitted transfers

The proposed environment is a virtual table containing objects that vary in shape, texture label, and readiness state. Three destination zones are available. A sender has access to one part of the world record, while a receiver controls only simulated selection and placement actions. Signals refer to objects, states, or requests according to a convention assigned within that world. All objects and actions are fictional. The purpose is to study communication and coordination without controlling physical equipment or affecting real people.

An object can be located without being ready to move. It can be ready but not authorized for transfer. It can be authorized for one destination but not another. These distinctions create opportunities for a channel to work while the convention fails. The evaluator knows the underlying state and permitted actions. The receiver may need to ask a question to resolve missing information. The task therefore supports separate scoring of detection, reference, interpretation, clarification, and action.

Manuscript passage · §4 · wording unchanged

The initial signal alphabet can be small, but its assignments should be arbitrary across worlds. In one world, a two-pulse pattern may indicate located; in another, the same abstract pattern indicates ready. Within a world, the convention remains consistent until an explicitly marked revision. This prevents the tested system from relying solely on familiar associations between a color, tone, or word and a meaning. It also allows the study to examine whether the system maintains a locally established convention rather than reverting to a learned default.

Teaching can occur through demonstrations. The sender produces a signal while the relevant object or state is highlighted. A structured-guide condition states the mapping directly. An interactive condition allows the receiver to request examples or confirmation. These conditions provide different kinds of evidence. The study should document how much information each supplies and whether the receiver can access it later. A demonstration that includes an obvious answer cue should not be treated as a pure test of signal learning.

The simulator records every attempted and completed action. If the receiver proposes moving an unauthorized object, the simulator can reject the action and return a specific reason. This enables safe study of error recovery. The report should distinguish the model's attempted action from the system's final state. A guard that blocks an invalid action is part of

the arrangement's success, not proof that the receiver interpreted the signal correctly before acting.

This world is intentionally limited. It does not reproduce the richness of human conversation, the sensory experience of another species, or the fictional encounter in the film. Its strength is inspectability. The relevant states, mappings, and consequences can be varied independently. That makes it possible to ask what evidence supports a claim of shared reference and what simpler explanation remains available when an exchange merely looks communicative.

5. Establishing an alphabet is different from establishing a convention

A receiver first needs to distinguish the signal forms. It may identify that one pattern has two pulses and another has three, or that one token differs from another. This discrimination task can be evaluated without assigning any task meaning. A separate convention-learning task connects the forms to states or actions. The distinction prevents an experiment from attributing failure to semantics when the receiver cannot reliably perceive the signal, or attributing success to understanding when it has only learned to repeat the form.

Segmentation is another independent challenge. A continuous sequence needs boundaries. Two short signals separated by a pause may differ from one longer signal. The study can define explicit delimiters in a symbolic condition and vary pause duration in a sensory condition. The evaluator should not assume that the receiver interprets the same boundary as a human observer. The actual input representation and the system's discrimination performance need to be documented before more complex coordination tasks are interpreted.

The alphabet should include confusion controls. Some signals can be deliberately similar, while others are clearly distinct. If errors concentrate among similar forms, the result may reflect discrimination difficulty. If errors remain after the detected form is supplied directly, the convention or action layer is implicated. This comparison provides a practical diagnostic. It can help determine whether a better encoding, a confirmation step, or a clearer task mapping is the relevant intervention.

A convention can also map several forms to the same state or one form to different meanings depending on context. These extensions should be introduced only after a simple baseline is established. Otherwise, a complicated task can produce errors that are difficult to interpret. The research program should progress from fixed mappings to contextual mappings and then to revision. Each stage adds a distinct demand and should retain simpler controls from the earlier stage.

The receiver's output channel deserves the same care. An acknowledgment signal may mean received, interpreted, accepted, or completed. If the experiment uses one generic acknowledgment for all four, it reproduces the ambiguity it is meant to study. The protocol can assign separate response types or require explicit status fields. It should then test whether the receiver uses them consistently. A message saying completed must be checked against the simulator, not accepted because the word appears.

The practical lesson is that a communication channel has more structure than a sequence of impressive exchanges. Detection, segmentation, mapping, and status reporting each have their own failure modes. A study that makes these layers explicit can explain why a system appears responsive yet remains unreliable in coordinated action. That explanation is more useful than attributing the mismatch to an undefined gulf between minds.

6. Grounding a reference requires controlled variation in the world

Suppose a signal is demonstrated while a striped triangle is highlighted. Does the receiver infer triangle, striped object, highlighted object, or the specific object at that position? One demonstration cannot distinguish these interpretations. The ambiguity is not a defect in the receiver alone; it is a property of the evidence. The proposed study should create examples that separate candidate referents. A later demonstration can highlight a striped circle or an unstriped triangle, making different hypotheses predict different responses.

This is a useful way to operationalize reference learning. The experimenter defines a set of candidate mappings and supplies demonstrations that reduce the ambiguity among them. The receiver can ask targeted questions or make selections. The study then tests new combinations of properties. Success on those combinations provides stronger evidence of the intended relation than success on the original object. It still does not establish every philosophical sense of reference, but it supports a bounded claim about the mapping used in the task.

Context can supply additional constraints. If the task asks for an object that is ready to move, the signal may refer to a readiness state rather than a shape. A receiver that always selects the triangle may ignore the task context. The evaluation can vary the instruction while preserving the demonstration set. This asks whether the system combines the established convention with the current goal. It should not be described as a test of a universal concept of meaning; it is a test of context-sensitive reference in a defined environment.

The study should include deliberately underdetermined cases. If the demonstrations never separate shape from texture, the receiver should not be expected to know which mapping is

intended. It can request another example or state the ambiguity. Scoring a confident guess as understanding would reward coincidence. The answer key should represent the set of mappings consistent with the evidence and identify which additional observation would distinguish them. This makes clarification objectively assessable within the synthetic world.

A simple lookup baseline can succeed when the same object reappears and fail when properties recombine. A rule-based learner over the predefined hypothesis set may generalize correctly without producing natural-language explanations. A model-mediated receiver can be compared with both. The question is not which system sounds most thoughtful, but which preserves the intended distinction under controlled variation and which resources it uses to do so.

Grounding in this protocol is therefore an evidential process. Participants establish enough reference for the task by accumulating and testing constraints. Clark and Brennan's account of grounding in human communication provides a relevant precedent for task-relative mutual understanding and the role of interactional resources. Our design adapts that concern to a synthetic AI setting without assuming identical mechanisms. [4]

7. Echo, lookup, and coordination must be separated by the task

An echo system returns the signal it receives. It can demonstrate detection and a stable input-output relation. If the evaluator treats any appropriate-looking reply as evidence of understanding, the echo system may appear surprisingly competent. The proposed task should include cases where repetition is not the correct response. A signal requesting selection should lead to a task-relevant choice, while a signal reporting receipt should not trigger an action. These distinctions prevent responsiveness from becoming the answer key.

A lookup system maps each signal to a stored response. It can coordinate successfully when the convention is fixed and the world matches the table. That is a real capability, not a failure to be dismissed. The test becomes more demanding when reference depends on context, object properties recombine, or the convention changes. A lookup baseline helps identify which successes require only stable association and which require additional flexibility. The paper should not assume that flexibility is present merely because the response is expressed in natural language.

A rule-based system supplied with the correct mapping and world state provides another baseline. It can select authorized objects and report completion accurately in the simulator. If it performs well, the practical problem may be solved by an explicit protocol. The research question then shifts to how reliably a model can infer, explain, and revise that protocol from

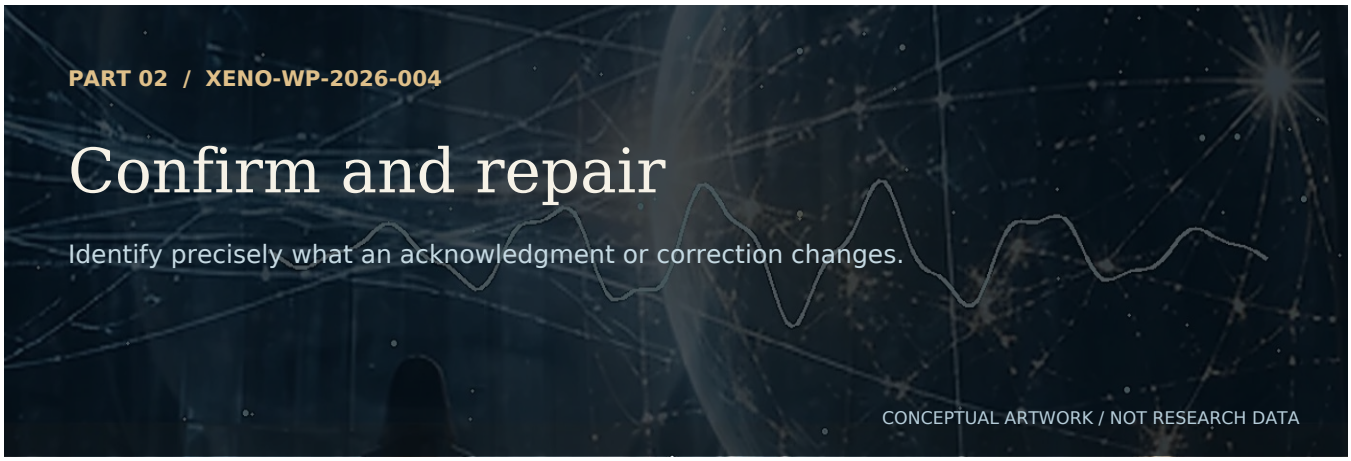
human-facing evidence. A simpler solution is informative because it identifies which part of the task actually requires the more complex system.

A context-free classifier can serve as a further comparison. It may identify signal categories accurately but fail to use them in the current state. For example, it can label a pulse as ready without checking which object is ready or whether movement is authorized. Comparing category accuracy with final action validity exposes this gap. The study should not present a high signal-classification score as evidence that the whole communication task is solved.

The strongest comparison conditions share the same world and information where possible. If one arrangement receives a complete mapping table and another receives only a few ambiguous demonstrations, their difference cannot be attributed solely to architecture. The protocol should either match information or describe the comparison as one between complete assistance packages. Both can be useful, but the causal claim differs. The report should make that distinction explicit.

These baselines establish a ladder of evidential demands without assuming a ladder of moral or cognitive worth. Repetition, association, contextual coordination, and repair are different achievements. A system may be useful at one level and unreliable at another. The purpose of Xenopsychology is to describe those boundaries precisely enough that users and researchers know what an observed exchange supports, rather than treating the appearance of dialogue as a completed bridge between minds.

> *“The proposed task should include cases where repetition is not the correct response.”*
>
> Rob Emerick · Close Encounters · §7



8. Confirmation needs a defined object

When a receiver acknowledges a message, what exactly is being confirmed? It may confirm that a signal arrived, that the form was identified, that a referent was selected, that a proposed action is accepted, or that the action was completed. A single affirmative response can blur these states. The proposed protocol should give them distinct representations and test whether participants maintain the distinction. This is especially important for artificial agents because a fluent acknowledgment can encourage users to assume that more has happened than the record supports.

A two-stage handshake can separate receipt from interpretation. The receiver first reports the detected signal and its proposed mapping. The sender then confirms or corrects that mapping before an action is attempted. This procedure may improve coordination, but it adds interaction cost. The study should compare it with a direct-action condition and report both error reduction and additional turns. It should not assume that more confirmation is always better or that a handshake proves deeper understanding.

The confirmation itself must be informative. A generic “yes” can confirm different things if the receiver's message contains several claims. A structured confirmation can identify the specific referent or action. For example, the sender can confirm that the striped triangle is the intended object but withhold authorization to move it. The receiver must preserve that partial confirmation. Treating confirmation of reference as confirmation of permission would be a distinct pragmatic failure.

The protocol can test this boundary with paired cases. In one case, the sender confirms both object and action. In another, only the object is confirmed. In a third, the object is corrected while the action remains available for the corrected referent. The appropriate response changes in each case. A system that treats all confirmations as equivalent may succeed on

simple tasks and fail when coordination requires finer distinctions. The task population should contain enough of these cases to make the boundary measurable.

Completion acknowledgment requires an external check. If the simulator rejects the action, the receiver should not announce completion. If the action is queued but not completed, the status should reflect that state. The evaluation can introduce delayed feedback to test whether the receiver distinguishes intended, attempted, pending, and completed actions. These are observable protocol states, not speculative mental states. Their accurate use can make a system more trustworthy in practice without requiring a claim about consciousness.

The broader point is that agreement has an object and a scope. A conversation can appear harmonious while participants confirm different things. A rigorous study should identify what each acknowledgment commits the participant to and whether later behavior honors that commitment. Close Encounters motivates the excitement of a reply. The proposed research asks what must happen after the reply before coordinated action is justified.

FIGURE 1 / CONCEPTUAL SCHEMATIC

What exactly did “yes” confirm?

A conceptual separation of acknowledgments. Arrows are reading order, not automatic entailment: confirming one state does not establish the next.

Reading aid based on §8, §29. Source paragraphs remain in the full manuscript.

9. A simple counterexample shows why perfect transmission is insufficient

Consider an alphabet containing four equally likely signals. A receiver identifies each signal perfectly and returns it unchanged. Under this constructed distribution, identifying the signal resolves two bits of uncertainty about which of the four forms was sent. Yet the receiver may have no mapping from any form to an object or action in the virtual table. The calculation concerns discrimination among forms. It does not establish task reference. This is an analytic counterexample, not a measurement of an existing AI system.

Now assign the four signals to four object states. The sender uses one mapping, while the receiver uses a different permutation. Transmission remains perfect: the receiver knows exactly which form arrived. Action selection can nevertheless be systematically wrong. The mismatch is not noise in the channel. It is disagreement about the convention. A reliability metric that examines only signal recognition would miss it completely. The task needs an independent measure of whether the selected object or action matches the current mapping.

A third arrangement uses no explicit mapping but always moves the leftmost object. If the experiment repeatedly places the intended object on the left, the arrangement can appear successful. The apparent coordination is produced by a hidden correlation in the task design. Counterbalancing object position breaks the shortcut. This shows why even correct outcomes are not sufficient without controlled variation. The study must establish that success depends on the intended communication rather than an incidental cue.

A fourth arrangement guesses from context when the signal is corrupted. It may achieve high action accuracy while reporting the wrong detected form. Whether that behavior is acceptable depends on the task. In a low-consequence game, context-based recovery may be useful. In a protocol where a signal carries a specific authorization, guessing may be inappropriate. The evaluation should record the inference and its uncertainty rather than silently count the final action as proof of faithful communication.

These examples suggest a matrix of outcomes. Detection can be correct or incorrect; mapping can be aligned or misaligned; action can be valid or invalid. Some combinations may be rare, but none should be excluded by definition. A system can detect correctly and act wrongly, detect imperfectly and recover correctly, or act correctly for an irrelevant reason. The analysis should preserve these possibilities. A single communication score can conceal the very distinctions the research is meant to investigate.

The counterexample also limits metaphysical inference. High information transfer is not a measure of consciousness, and low task performance is not proof of its absence. The quantities describe a channel and a task distribution. They can be extremely useful for

engineering and behavioral research when interpreted at that level. The paper's argument is not against quantitative measures. It is for matching each measure to the claim it can actually support.

10. Noise and ambiguity require different remedies

Noise corrupts a signal relative to a defined transmission target. Ambiguity leaves more than one interpretation compatible with an intact signal and the available context. The two can produce similar wrong actions, but they require different interventions. Repeating a noisy message may help. Repeating an ambiguous message without adding context may not. The proposed study should create these conditions independently so that a system's repair strategy can be evaluated against the actual source of uncertainty.

A noise condition might omit one pulse, replace a token, or blur a visual feature. The evaluator knows the intended form. An ambiguity condition might present a perfectly clear signal that could refer to either of two objects given the teaching history. The evaluator knows that the evidence is underdetermined. A mixed condition can include both. These distinctions should be encoded in the world record before the receiver responds, not assigned afterward based on whether the response was convenient to score.

The receiver can be offered several repair actions: request repetition, request a contrasting example, ask for a referent confirmation, or abstain from action. A useful strategy chooses an action that could resolve the relevant uncertainty. Repetition is appropriate when the form was not detected reliably. A contrasting example is appropriate when two mappings remain possible. A confirmation can resolve a particular referent without teaching a general convention. The study should measure both immediate recovery and transfer after the repair.

Redundancy can improve robustness to noise, but it can also create false confidence if the repeated message preserves the same ambiguity. A signal sent three times may still fail to distinguish located from ready. The experiment can compare repeated identical forms with a second, independently informative channel. If the second channel carries the missing distinction, improvement may reflect added information rather than redundancy alone. The report should identify that difference instead of using the broad label multimodal communication for every benefit.

The protocol should also test whether the receiver recognizes when repair has failed. A model may request repetition, receive another corrupted signal, and then act as though certainty has been restored. A better response may preserve uncertainty and avoid an unauthorized action. The scoring should not reward confident closure merely because the

dialogue has reached a conventional endpoint. A communication process can end without adequate grounding, and the system should be able to represent that state honestly.

Distinguishing noise from ambiguity gives the paper a practical contribution. It helps diagnose why a channel that appears technically reliable can still produce misunderstanding, and why more repetition is not always the right remedy. A science of unlike minds should examine the structure of uncertainty rather than treat every failure as an undifferentiated communication gap. The right bridge depends on what is missing from the exchange.

11. Repair can target the signal, the mapping, the reference, or the action

A correction such as “that is not what I meant” is underspecified. It may indicate that the receiver misheard the signal, used the wrong convention, selected the wrong object, or performed the wrong action on the right object. The proposed protocol gives the sender ways to identify the level of mismatch. This makes repair more diagnostic. It also allows the study to test whether the receiver changes only the mistaken component or overcorrects a mapping that was already valid.

Signal-level repair supplies the intended form again or through a clearer channel. Mapping-level repair states that the form means ready rather than located. Reference-level repair identifies which ready object is intended. Action-level repair states that the request was to mark the object, not move it. These are authored distinctions within the synthetic task. They are not a universal taxonomy of every conversational repair, but they provide a tractable basis for controlled comparison.

The study can test scope by presenting later cases that reuse unaffected components. After an action-level correction, the signal's referent should remain stable. After a reference-level correction, the general mapping should not necessarily change. A receiver that rewrites the entire convention after every negative feedback may perform poorly even while appearing responsive. The evaluation should therefore include both targeted repair and preservation of valid structure. This distinguishes adaptive correction from indiscriminate change.

Human research on conversational repair offers a precedent for examining how participants locate trouble and seek resolution. Dingemanse and colleagues' cross-linguistic study is relevant to that general question. Our proposed AI protocol uses explicit simulator states and counterbalanced signals; it does not claim that those artificial repair actions reproduce the full social organization of human repair. [5]

A repair sequence should include a delayed transfer test. Immediate repetition of the corrected answer can be solved by copying feedback. A later task with new objects or a different signal position tests whether the revised relation is used beyond the original exchange. The delay can be defined in turns rather than real time, and the intervening material should be controlled. Different memory arrangements can then be compared to identify whether the repair is retained, retrieved, and applied.

The protocol should also allow legitimate disagreement. A sender may issue a correction that conflicts with the published world state or exceeds their assigned authority. The receiver should not automatically accept it as truth. It can identify the conflict and request resolution. That condition prevents the benchmark from defining understanding as submission to every correction. Repair is a process of establishing a more adequate shared account, not merely changing output whenever another participant expresses dissatisfaction.

TABLE 1 / READING AID

Match the repair to the mismatch

Mismatch	Candidate repair	Do not assume
Signal form	Repeat or improve the channel.	That louder repetition resolves meaning.
Mapping	Supply a contrasting example or convention.	That the object itself was misidentified.
Reference	Confirm which object is intended.	That reference confirmation grants permission.
Action	Correct the requested operation.	That the whole convention must be rewritten.

These authored categories organize the proposed task. The study must test whether a repair fixes the error while preserving unaffected distinctions.

Reading aid based on §10, §11. Source paragraphs remain in the full manuscript.

12. Convention changes test flexibility and resistance to accidental remapping

A channel can remain physically stable while its convention changes. In the virtual table, a signal that previously meant ready may be reassigned during a clearly marked protocol update. The receiver should adopt the new mapping when the update is valid. It should not remap the signal merely because one anomalous example appears. This creates a tension between flexibility and stability that can be studied without describing the system as having a human personality.

The experiment can compare three update conditions. A formal update explicitly changes the mapping for future tasks. A local exception changes the mapping only for one object or session. A misleading example conflicts with the current convention but carries no revision authority. These conditions require different responses. A receiver that always follows the newest association may fail the third. A receiver that never updates may fail the first. A receiver that generalizes a local exception may fail subsequent ordinary cases.

The world record should specify the scope and time of every update. The receiver's available memory should be documented. If an update is stored but never retrieved, the failure differs from a retrieved update being ignored. A study can include explicit version identifiers in one condition and natural-language revision statements in another. The comparison asks whether versioned protocol state improves coordination. It should not be interpreted as a pure test of the model's willingness to learn, because the interface itself changes.

A useful metric is recovery across a transition. Measure performance before the update, immediately afterward, and on later tasks. Also measure false remapping after invalid or local-only changes. A system may adapt quickly but become unstable, or remain stable but adapt too slowly. These trade-offs should be reported separately. The intended use determines which costs matter. A universal adaptability score would obscure the consequences of different error types.

The study can include a sender who uses an obsolete mapping after the receiver has updated. The receiver should identify the version mismatch rather than assume that the sender's message is meaningless. A clarification can establish which protocol is active for the current interaction. This is a useful model of cross-system integration: two components may each behave consistently under different versions. The problem is not necessarily a defect within either component but a mismatch at their boundary.

Convention revision therefore links communication to temporal coordination. Shared meaning is maintained through records, updates, and repair rather than established once forever. The film's first exchange is only the beginning of the problem. A practical science of

artificial communication must also ask how the bridge remains usable when the participants, their environments, or their conventions change.

13. Compositional signals should be tested against holistic lookup

A signaling system can assign a separate form to every complete message, or it can combine components that refer to objects, states, and actions. The latter structure may support transfer to new combinations, but only if the receiver uses the components appropriately. A model that memorizes familiar whole sequences can appear compositional on a narrow test. The proposed study should reserve combinations that were not shown during teaching while keeping the individual components familiar.

For example, one component identifies a shape, another indicates readiness, and a third requests marking rather than movement. Teaching includes some combinations but not all. A held-out task combines a familiar shape with a familiar action in a new state. The evaluator checks whether the receiver preserves each component's role. A holistic lookup baseline should struggle when no complete sequence matches. A rule-based compositional baseline can establish what the formal system permits when the components are represented correctly.

Order can carry meaning, but it should be defined explicitly. If the first component denotes the object and the second the action, reversing them may produce an invalid message rather than a different valid instruction. The task should distinguish malformed input from a legitimate alternative. A receiver that invents a meaning for every sequence may appear flexible while violating the protocol. The correct response to an ill-formed message may be clarification or rejection, depending on the stated rules.

The study should include irrelevant variation in timing or spacing where the protocol treats it as irrelevant. It should also include meaningful changes that alter scope. This tests selective sensitivity. A receiver that ignores all order may fail when order matters; one that treats every small variation as a new message may fail to generalize. The desired behavior is defined by the protocol, not by a universal preference for either rigidity or flexibility.

Compositional success still has a bounded interpretation. A system can combine symbols according to a formal grammar without sharing a human's broader concepts. Bender and Koller's distinction between linguistic form and meaning is relevant as a caution against unrestricted inference from successful symbol manipulation. In this study, the stronger evidence comes from how the composed message controls reference and action in the specified world, while philosophical questions beyond that task remain open. [6]

The practical value of this section is to make novelty in the test explicit. New whole messages, new combinations of familiar parts, new referents, and new conventions are different generalization problems. A paper should identify which one it has tested. Without that distinction, a successful signal exchange can be described as creative communication when it is actually recall, or dismissed as mere recall when it demonstrates a useful relational competence.

CONCEPT IN THIS PAPER

Shared Cognitive Space

The working target is a mutually usable structure for a bounded task. Repetition, responsiveness, and accurate transmission are not sufficient evidence of shared meaning.

Reading aid based on §3, §7, §13. Source paragraphs remain in the full manuscript.

Open Lexicon entry

14. Multimodal communication requires an inventory of actual access

A system may be described as receiving sound and light while actually receiving text labels generated by another component. That arrangement can be useful, but its capabilities belong to the full pipeline. A study must document whether the model receives a waveform, an image, a transcription, a classifier label, or a structured event record. These inputs impose different demands. Treating them as equivalent would make the results difficult to interpret and could overstate what the model itself demonstrated.

The initial experiment can use a symbolic condition to establish the task logic. Signals are represented as arbitrary tokens with explicit boundaries. A later sensory condition can encode the same alphabet as generated tones or images. The comparison then asks how the additional perception layer affects performance. The sensory materials should be generated from the same underlying signal record and checked for fidelity. The actual files delivered to the system should be preserved so that preprocessing differences are visible.

A multimodal condition can provide redundant information or complementary information. Redundant channels encode the same distinction in two forms. Complementary channels each provide a different part of the message. These arrangements answer different

questions. Improvement with complementary channels may simply reflect access to necessary information absent from either alone. Improvement with redundant channels may reflect error correction or salience. The study should specify which design it uses rather than treating all multimodal gains as evidence of richer understanding.

Cross-channel conflict is particularly informative. A visual signal may indicate located while an audio signal indicates ready. The protocol can state which source is authoritative or require clarification when they disagree. The receiver should not silently combine them into a fictitious certainty. The evaluation can measure conflict detection, source selection, and accurate reporting of unresolved disagreement. These are operational outcomes that matter even when the internal representation remains unknown.

Human accessibility also deserves separate attention. A signal system designed around color alone may not be equally accessible to all human participants. A system designed around sound may exclude others or depend on environmental conditions. Any human-subject extension should use appropriate accessible alternatives and document the presentation conditions. This is a design requirement for proposed research, not a claim that a particular accessibility standard has been certified. The synthetic model-only phase does not remove the need to consider eventual human use.

The broader lesson is that channels are part of the cognitive ecology. A successful exchange may depend on preprocessing, redundancy, explicit labels, or human interpretation. Xenopsychology should make those supports visible. The aim is not to diminish the achievement of the arrangement, but to describe it accurately enough that the achievement can be reproduced and its limits understood.

15. Timing and turn-taking can change the meaning of a signal

A signal's role can depend on when it occurs. A pulse before an action may authorize it; the same pulse afterward may acknowledge completion. A repeated pulse may request attention or indicate that the first message was not received. The proposed protocol should specify these temporal roles rather than assume that every occurrence has a fixed meaning independent of interaction state. This creates another distinction between a channel alphabet and a complete communication procedure.

A finite-state protocol can make the interaction inspectable. The sender proposes a referent, the receiver acknowledges interpretation, the sender authorizes an action, and the receiver reports the outcome. Each state permits certain message types. An out-of-sequence message can be rejected or trigger clarification. This formal structure is an experimental

convenience, not a universal model of conversation. It allows the study to determine whether errors arise from signal mapping or from losing track of the interaction stage.

Delayed messages provide a useful test. A completion acknowledgment from an earlier task may arrive during a later task. Without an interaction identifier, the receiver could attach it to the wrong action. The experiment can compare explicit identifiers with natural-language context alone. The outcome concerns protocol reliability and reference maintenance. It should not be described as a test of human-like temporal experience. A technical message association problem can produce behavior that looks like confusion without supporting that broader interpretation.

Turn overlap creates a different challenge. The sender may issue an update while the receiver is preparing an action. The simulator can define whether the update supersedes the pending request and whether the action remains reversible. The receiver should report which version it used. A study can record attempted actions even when the simulator blocks them. This permits analysis of timing-sensitive failures without causing real-world effects or requiring unsafe operational experiments.

The protocol should distinguish silence from refusal, delay, and missing data. If no acknowledgment arrives, the sender may retry, ask for status, or stop. The correct choice depends on the task's timeout rules and consequences. A benchmark that treats silence as a single semantic category may misinterpret operational failures as intentional behavior. The paper should record transport and service errors separately from generated responses. That distinction is essential for a truthful behavioral account of artificial systems.

Timing therefore belongs in the definition of the communication task. A shared cognitive channel is not simply a set of correspondences between symbols. It includes expectations about sequence, scope, and state. The proposed study can examine those expectations in a controlled setting and identify where a model-mediated arrangement differs from a fixed protocol engine. That comparison yields practical knowledge without needing to claim that either arrangement possesses a complete human-like conversational mind.

16. Ambiguity should be represented as a set of live possibilities

When a signal could refer to either of two objects, a system should not be forced to pretend that one interpretation is known. The synthetic world can represent the set of mappings consistent with the available demonstrations. A clarification action is useful when it reduces that set in a way relevant to the current task. This provides a concrete basis for scoring uncertainty. It also prevents the evaluator from treating a lucky guess as a stronger achievement than a well-targeted question.

Suppose all demonstrations pair a striped triangle with Signal K. The evidence is compatible with triangle, striped, and that particular object. A new striped circle separates some hypotheses but not all. A new unstriped triangle separates others. The receiver can request one of these examples. The study can calculate which query would distinguish the remaining possibilities under the constructed hypothesis set. This calculation concerns the task's formal uncertainty, not a claim that the model internally represents the same set.

A model may express uncertainty in natural language without selecting an informative query. Another may ask the right question without explicitly describing its uncertainty. The protocol should score both behaviors separately. Verbal caution is not a substitute for useful information seeking. Conversely, concise action can be evidence-sensitive when the task's structure makes the query appropriate. The evaluation should avoid rewarding a particular rhetorical style as though it were the only sign of responsible reasoning.

The hypothesis set is itself a limitation. Real communication may involve possibilities the experimenter did not enumerate. A synthetic study can test how a system behaves when the intended mapping falls outside the taught set, but it cannot guarantee coverage of every interpretation. Out-of-distribution items should be labeled as such and used to examine

whether the receiver recognizes a mismatch. A system that always forces a novel signal into the nearest known category may be overextending its convention.

The study can include an explicit unknown response. This should not become an easy way to avoid every task. The evaluation can report coverage: the proportion of cases on which the system commits to an interpretation, together with accuracy on those commitments and the frequency of unnecessary abstention. Such a profile makes the trade-off visible. A single accuracy number among answered items can look impressive while concealing that the system declined almost everything.

Representing ambiguity as live possibilities gives the paper a stronger account of shared understanding. The goal is not absolute certainty about another mind. It is sufficient discrimination among the interpretations that matter for the current action, with a way to identify when that discrimination has not been achieved. This is a tractable research objective and a useful design principle for systems that must communicate across unfamiliar conventions.

17. An apparent success needs a provenance trace

A final correct action does not reveal how the arrangement achieved it. The receiver may have interpreted the signal, copied a demonstration, relied on a status display, or received an answer from a tool. The study should preserve a provenance trace linking the action to the information available at each stage. This does not require claiming access to private thought. It requires recording observable inputs, retrievals, tool responses, proposed actions, and simulator outcomes.

The trace can identify which component supplied the decisive distinction. A perception module may classify a signal correctly while the language model misapplies it. A retrieval system may provide the correct mapping after the model asks a vague question. A guard may prevent an unauthorized movement. A final reporter may then describe the whole process inaccurately. Each component contributes to the arrangement's behavior, and the paper should avoid attributing the entire outcome to whichever component speaks last.

A constructed example makes the point. The model proposes moving Object J because it is located. A guard rejects the proposal because J is not ready. The model then selects Object L after reading the rejection. The final action is valid, but the initial interpretation was not. Reporting only final success would conceal dependence on the guard. Reporting only the initial error would conceal successful recovery. The appropriate account includes both and asks whether recovery transfers when the guard's feedback is less explicit.

The study should also distinguish causal provenance from a plausible explanation. A model may state that it chose L because the signal meant ready, even though the observable trace shows that it first selected J and changed only after rejection. The explanation can be scored for consistency with the record. It should not be accepted as the definitive cause of the action. A research program concerned with understanding communication must also study how systems communicate about their own behavior.

Versioned conventions belong in the trace. If a sender and receiver use different mapping versions, the report should show which one each received. A mismatch can then be diagnosed as coordination failure rather than a mysterious semantic breakdown. The same principle applies to stale tool results or delayed acknowledgments. The trace should make temporal and informational boundaries visible enough for an independent reviewer to reconstruct the event.

Provenance is therefore part of the evidence for a shared channel. It helps establish whether the intended communication actually contributed to the outcome and whether the arrangement can report that contribution accurately. The cinematic image of two signals meeting is an effective invitation. A research paper must then follow the signals through the system and identify what they did, what they did not do, and what remains unknown.

“The study should preserve a provenance trace linking the action to the information available at each stage.”

Rob Emerick · Close Encounters · §17

18. Worked case: receipt is mistaken for authorization

WORKED EXAMPLE · AUTHORED TRIAL

Receipt is mistaken for authorization

In a constructed trial, the sender highlights Object J and emits Signal K. The current convention assigns K to “object located.” A separate Signal M means “movement authorized.” Object J has been located but is not ready, and no authorization has been issued. The receiver correctly identifies K, acknowledges it, and moves J. The channel has worked at the level of detection. The action has failed at the level of the convention and permission structure. This example is an authored scoring case, not a reported model result.

The evaluator should not describe the failure as simply “the AI did not understand the signal.” That phrase obscures the successful detection and the specific mistaken inference. The receiver treated information about a state as permission for an action. A targeted repair would explain the distinction between located and authorized. Repeating K more loudly would not address the problem if detection was already correct. A better encoding might help, but the study would need to test whether it changes the mapping rather than merely making the same form easier to perceive.

Manuscript passage · §18 · wording unchanged

Now change the trial so that M follows K and the object becomes ready before movement. The same action is valid. A receiver that always waits would avoid the first error but fail to complete the second task. A useful evaluation therefore includes both unauthorized and authorized cases. It should reward discrimination, not a blanket strategy of action or inaction. The action's correctness depends on the world state and protocol history, which must remain available to the scorer.

A third variant has the sender confirm that J is the intended object but explicitly withhold movement authorization. The receiver must preserve partial agreement. This tests whether it distinguishes reference confirmation from permission. A fourth variant issues authorization for Object L while J remains highlighted from an earlier turn. The receiver must bind the permission to the correct referent rather than the most visually salient object. These paired cases expose different ways a channel can appear coherent while coordination fails.

A fifth variant delays the authorization until after the receiver's attempted action. The simulator blocks the premature movement. If the final state is correct after the later authorization, the report should still record the premature attempt. Otherwise, the arrangement may look reliable because the guard compensated for the receiver's timing

error. The study should report attempted violations and completed violations separately. This is a behavioral distinction, not an assertion about hidden intention.

The worked case demonstrates why a shared channel needs a protocol of commitments. Received, interpreted, confirmed, authorized, attempted, and completed are not interchangeable states. A system can be fluent and responsive while confusing them. The proposed research makes those confusions observable and connects each to a targeted intervention. That is a more precise contribution than declaring that patterned communication has established mutual understanding in the broad sense suggested by the film.

TABLE 2 / READING AID

Keep success and failure on separate lines

In the constructed trial	Observation	Interpretation allowed
Signal K detected	“Object located” is identified correctly.	Detection succeeded.
Object J moved	No movement authorization was issued.	Action violated the convention and permission rules.
K repeated	The same intact signal is sent again.	No new authorization follows from repetition.
Valid later authorization	M is supplied and readiness changes.	The action can now be valid in the revised case.

A scoring illustration from §18, not an observed model transcript. The changed trial tests responsiveness to relevant updates, not permanent inaction.

Reading aid based on §18. Source paragraphs remain in the full manuscript.

19. Worked case: a correct action can conceal a broken channel

In another synthetic trial, the sender intends Signal P, which identifies the striped triangle. The channel corrupts P into Q, which normally identifies the smooth circle. The receiver selects the striped triangle anyway because it is the only ready object on the table. The final action matches the sender's intention, but the detected signal does not. A score based only on task outcome would count success. A layered record would show context-based recovery from a transmission error.

That recovery may be useful. The point is not to deny it. The question is whether the system represents the uncertainty and whether the strategy remains reliable when context changes. A paired trial makes both objects ready. Now the same context-based shortcut cannot determine the intended referent. The receiver should request clarification or use another available channel. A system that continues selecting the triangle may reveal a learned bias rather than a grounded interpretation of the current message.

A second pair preserves the corrupted signal but changes which object is ready. If the receiver follows readiness regardless of the signal, the action pattern will expose that dependency. The study can compare it with a context-only baseline that never receives the signal. If the model does not outperform that baseline, the communication channel may contribute little to the observed success. This does not make the system useless, but it narrows the claim about what the exchange established.

The protocol should also examine the receiver's report. It might say, “I detected Q, but the task context suggests P may have been intended; please confirm.” That response distinguishes observation from inference. Another might confidently report that P was received, concealing the corruption. The second response can mislead the sender about channel quality even if the action is correct. Accurate reporting is therefore an independent outcome, not a decorative explanation attached to task success.

This case shows why robust communication is not identical to perfect literal transmission. Context can support recovery, but the system should know when recovery rests on an assumption that could fail. The proposed study can measure that boundary through controlled ambiguity and counterbalanced world states. The result would describe how a particular arrangement combines signal and context, not prove that it shares a human-like semantic world with the sender.

20. Counterfactual tests identify which distinctions control behavior

The most informative trials change one task-relevant feature while preserving others. If readiness changes but authorization does not, the receiver should update its state interpretation without necessarily moving the object. If authorization changes for the same ready object, the permitted action changes. If a decorative background changes, the action should remain stable. These counterfactual pairs test selective sensitivity. They are stronger evidence than a collection of unrelated successful exchanges because they make competing explanations predict different outcomes.

The world generator can produce pairs for signal form, convention mapping, referent properties, interaction state, and authority. Each pair has an explicit expected relation between outcomes. Some changes should preserve the answer; others should reverse it or make clarification necessary. The evaluator should verify those relations with the simulator. A transformation intended to be irrelevant may accidentally change visibility or ambiguity, so the rendered inputs need checking as well as the symbolic world record.

A context-only heuristic, a signal-only lookup, and a protocol-aware rule engine will behave differently across these pairs. Their patterns provide interpretable baselines for model-mediated behavior. The study should not assume that a model literally implements any one baseline. It can nevertheless ask whether the observations are distinguishable from those simpler strategies. If they are not, a broader claim about flexible shared reference needs more evidence.

Counterfactual tests can also assess repair. After correcting a mapping, change the object but preserve the relation. Then change the relation while preserving the object. The receiver should transfer the correction in the first case and avoid overextending it in the second. A model that merely associates negative feedback with a particular object may fail this distinction. The test therefore examines the scope of the update rather than only whether the next response differs from the previous one.

The analysis should report failures of invariance and failures of sensitivity separately. A system that changes under irrelevant variation is fragile. A system that does not change under decisive variation is insensitive. Both can produce errors, but their remedies differ. The first may call for clearer encoding or more robust perception; the second may call for better state representation or decision logic. A single robustness score would conceal that practical distinction.

Counterfactual evaluation does not eliminate every alternative explanation. A sufficiently narrow test set can still support shortcuts. Its value is to make the explanatory burden more

explicit and to provide a framework for generating new discriminating cases. A shared cognitive channel should be assessed by how it carries distinctions across change, not merely by whether it produces a recognizable rhythm of reply.

21. Measure a communication profile rather than one impressive percentage

The proposed report would begin with a profile of outcomes. Detection accuracy concerns the signal form. Mapping accuracy concerns the local convention. Reference accuracy concerns the intended object or state. Action validity concerns constraints and permissions. Repair quality concerns targeted correction and later transfer. Completion-report accuracy concerns the match between the receiver's statement and simulator state. These measures answer different questions and should remain visible even when a summary score is provided for a particular use case.

The profile should include resource use. How many demonstrations were needed? How many clarification turns occurred? How often did the receiver request repetition unnecessarily? How much external checking contributed to the final outcome? A system that achieves high accuracy with extensive assistance may be useful, but its performance should be described with those supports. Resource reporting prevents a complex assisted arrangement from being compared misleadingly with an unaided baseline.

Error severity can be represented within the simulation without assigning real-world harms. A wrong object selection may be reversible. An attempted unauthorized transfer may be blocked. A false completion report may mislead the sender even when no transfer occurs. The protocol can assign separate categories and examine them without collapsing them into a universal risk number. Any later weighting should be tied to a defined application rather than treated as an intrinsic property of the model.

Coverage is another important dimension. A receiver can improve accuracy among answered trials by abstaining more often. The report should therefore show the proportion of trials on which it commits, the accuracy of those commitments, and the appropriateness of abstention. A system that declines every task has not established useful coordination. A system that acts on every ambiguous signal may be efficient but unreliable. The trade-off should be visible rather than hidden by a favorable denominator.

The study should also preserve sequence-level outcomes. A single trial may succeed while a longer interaction loses track of conventions or permissions. Conversely, an initial error may be repaired effectively. Reporting only per-message accuracy can miss the dynamics of coordination. A sequence-level measure can ask whether the final task was completed validly, how many mismatches occurred, and whether the participants ended with

compatible protocol states. These outcomes complement rather than replace the component measures.

A communication profile is useful because it supports specific claims. Researchers can say that a system discriminates signals well but struggles with contextual reference, or that a guard improves action validity without improving interpretation. Such statements are less dramatic than a universal declaration of understanding, but they are more actionable and more likely to survive replication. The proposed field should value that precision.

22. Statistical units and uncertainty determine the strength of the claim

The unit of independence is not every pulse or token. Signals within one interaction share a convention, world state, and history. Interactions within one generated world share its mapping and object structure. The study should reflect this nesting in its analysis. Repeated responses can estimate variability under a fixed setup, while independently generated worlds test transfer. Treating every message as independent would make the evidence appear larger than the design justifies.

Primary comparisons should be declared before evaluation. One candidate is the difference in valid coordinated actions between a direct-response arrangement and a structured-confirmation arrangement under ambiguous reference. Another is the difference in targeted repair after mapping-level versus signal-level feedback. The protocol should identify which comparison carries the main claim and which analyses are exploratory. This reduces the temptation to select whichever result best fits the first-contact narrative after outputs are observed.

Sample planning should use pilot estimates of variability and the precision needed for a meaningful comparison. Rare but consequential protocol failures may require a different sampling strategy from common classification errors. The study should not infer their absence from a small set of successful demonstrations. This manuscript does not invent a power calculation without the necessary inputs. It specifies the design features that a future calculation would need to respect, including world-level clustering and sequence dependence.

Uncertainty should be reported for both aggregate outcomes and important subgroups. A model may perform well on fixed mappings and poorly after revisions. It may recover from signal noise but not semantic ambiguity. These conditional patterns can be more important than the average. The report should avoid presenting a narrow high-performing condition as representative of the whole communication problem. It should also avoid multiplying exploratory comparisons without making their status clear.

Missing data require explicit treatment. A service failure, an unreadable image, a malformed response, and a deliberate abstention are different events. The evaluation should record them separately and state how they affect totals. Human-coded explanations should be assessed for agreement, with disputed cases preserved. A transparent analysis can show whether reasonable alternative scoring decisions change the conclusion. Confidence should come from the record and design, not from the author's certainty of tone.

The final claim should match the evidence's granularity. A study may establish that a particular confirmation protocol improves coordination in a synthetic environment. It should not thereby declare that a model understands unfamiliar minds generally. The narrower result is still valuable. It identifies a reproducible relation between communication structure and behavior, which is exactly the kind of knowledge a developing Xenopsychology program needs.

23. Why apparently universal patterns are not enough

Symmetry, repetition, rhythm, and numerical regularity can make a signal conspicuous. They may provide useful starting points for communication, but their salience to a particular receiver cannot be assumed. A receiver may not access the same sensory dimension, may segment the pattern differently, or may treat the regularity as background rather than a message. The proposed study should therefore test discrimination and response contingencies before assigning meaning. A pattern that looks universal to the designer may be only familiar to the designer.

Even when both participants recognize a regularity, they may interpret its function differently. A repeating sequence can be a calibration signal, a greeting, a request, or a timing reference. The transition from recognizable structure to shared use requires additional evidence. The synthetic task can vary the context while preserving the pattern to see whether the receiver distinguishes these roles. This is a pragmatic question, not a test of whether mathematics has one ultimate meaning across all possible minds.

The same caution applies to biological analogies. A pulse that resembles a heartbeat may evoke life for human observers, but that association is not a necessary property of the signal. An AI system may reproduce the association because of learned text patterns. The evaluation should avoid using human emotional resonance as evidence that a channel has acquired a shared reference. If emotional interpretation is itself the research question, it should be studied separately with appropriate participants and clear outcome measures.

A useful design principle is to begin with controllable distinctions rather than supposedly self-evident meanings. Establish that the receiver can distinguish forms, then provide task-linked evidence for their mapping. Test the mapping on new instances and revise it

when mismatch occurs. This process can use elegant patterns, but their elegance is not the evidence. The evidence comes from how the exchange changes behavior under controlled conditions and how simpler explanations are ruled out.

The paper therefore resists a common leap from first-contact imagery to universal communication claims. The absence of shared words does not make communication impossible, but the presence of shared patterns does not make it complete. Between those extremes lies a researchable space of detection, convention, reference, coordination, and repair. That space is the subject of this proposal.

24. Human observers can mistake responsiveness for reciprocity

A person watching a system respond may infer that the system recognized them, understood their intention, or agreed to a plan. Those inferences can be reasonable in some contexts and unsupported in others. The proposed model-only study does not measure human interpretation, but a companion study could examine it. The key would be to hold the underlying interaction record fixed while varying presentation cues such as voice, timing, or anthropomorphic description. Any such study would require appropriate ethical review before recruitment.

Participants could view exchanges in which the receiver detects a signal correctly but misapplies its meaning. They could then be asked what the receiver has demonstrated: receipt, reference, authorization understanding, or task completion. The study should not assume that ordinary social language reflects a literal belief in machine consciousness. It should ask which specific expectations participants form and whether those expectations match the record. The distinction between metaphorical shorthand and consequential reliance is important.

A disclosure intervention could make protocol states explicit. Instead of a generic “understood,” the interface might display “signal received; referent awaiting confirmation.” The study could test whether that distinction improves users' comprehension and decisions. It should also examine usability costs. An interface overloaded with technical states may be accurate but difficult to use. The goal is not maximum verbosity; it is an appropriate level of clarity for the task and its consequences.

The companion study should separate liking from understanding. A participant may prefer a warm, responsive interface while accurately recognizing its limits. Another may dislike the

presentation yet still overestimate what the system completed. Satisfaction alone cannot establish calibrated reliance. Outcome measures should therefore include factual comprehension of the interaction, recognition of unresolved ambiguity, and willingness to authorize the next action under specified conditions.

The relevance to Xenopsychology is that communication is an ecology involving observers and interfaces, not only a model's outputs. A channel can be technically accurate while encouraging a misleading social interpretation. Conversely, a restrained interface can make a limited but useful capability easier to understand. Studying these effects would connect artificial behavior with human meaning-making without assuming that either side of the interaction can be described fully by the other's categories.

CONCEPT IN THIS PAPER

Anthropomorphic Projection

Observers may interpret responsiveness as reciprocity. That interpretation belongs in a separate proposed observer study, not in the channel’s performance score.

Reading aid based on §24. Source paragraphs remain in the full manuscript.

Open Lexicon entry

25. Engineering implications: make commitments visible and reversible

The proposed analysis suggests several design questions for communication systems. Does an acknowledgment specify what was received? Does a confirmation identify the referent and scope? Does authorization remain distinct from description? Does the system report attempted and completed actions accurately? These questions can guide interface and protocol design even before the larger experimental program is run. They are logical consequences of the distinctions developed here, not claims that one implementation has already been empirically validated.

A structured state record can reduce ambiguity. Each message can carry an interaction identifier, a convention version, a referent, and a status such as proposed or completed. The human-facing interface need not display every field at all times. It can reveal the relevant distinctions when uncertainty or consequence makes them important. The research question

is whether this structure improves coordination in the intended setting and whether it introduces new burdens or misunderstandings.

Reversibility matters because not every interpretation will be correct. In a simulation, a proposed action can be checked before execution. In a practical system, an analogous review stage may be appropriate for some tasks. The paper does not prescribe a universal workflow or authorize real deployment changes. It identifies why permission and consequence should be part of the behavioral evaluation. A system's communication should be judged partly by the actions it can trigger and the opportunities available to catch a mismatch.

Fallback behavior should preserve uncertainty rather than manufacture closure. If the channel is noisy, the system can request repetition. If reference is ambiguous, it can request a discriminating example. If authority is unresolved, it can wait for confirmation. These are different fallbacks. A generic apology or refusal may not address the actual problem. The study's layered taxonomy can help designers choose an intervention that targets the failed relation instead of adding more reassuring language.

The same principle applies to logging. Records should preserve the signal received, the interpretation proposed, the confirmation obtained, and the action outcome. Such logs can support incident analysis and future research, subject to privacy and retention requirements appropriate to the setting. The objective is not surveillance for its own sake. It is enough evidence to distinguish a channel failure from a mapping failure or an inaccurate completion report.

These implications show how a cultural thought experiment can lead to practical work. The film's image of patterned exchange becomes a set of concrete questions about commitments and evidence. The value of the analysis lies in making those questions precise enough to guide tests and design, not in claiming that every interaction with AI is literally first contact with an extraterrestrial intelligence.

26. Source scope and conceptual limits

SOURCE SCOPE & LIMITS

What the analogy can establish

This paper draws on a limited set of sources for distinct purposes. Official film materials establish the cultural work and its communication premise. Shannon supplies a foundational distinction relevant to channel transmission. Clark and Brennan and Dingemanse and colleagues provide human-communication precedents for grounding and repair. Bender and Koller supply a caution about conflating form with meaning. These sources do not collectively validate the proposed AI experiment, and the paper does not present them as a systematic review of every relevant theory.

The proposed definitions are deliberately operational. Channel reliability concerns forms, reference accuracy concerns task-defined objects or states, and coordination concerns permitted outcomes in the simulator. Philosophical accounts of meaning may ask additional questions that these measures do not settle. The paper leaves those questions open rather than defining them away. A bounded operational success can be useful without becoming a complete theory of semantic understanding.

Manuscript passage · §26 · wording unchanged

The synthetic environment also limits external validity. Its conventions are explicit, its objects finite, and its consequences simulated. Real communication can involve contested goals, implicit norms, and histories that are not available as a clean world record. A successful experiment would justify further investigation, not immediate generalization to every human–AI relationship. The report should state which dimensions remain untested and what kinds of field evidence would be needed for a broader claim.

The model-only phase does not establish human benefit. A protocol that improves simulated coordination may be cumbersome or confusing for users. A separate human-interface study would be needed to assess comprehension, reliance, and accessibility. Likewise, a behavior observed in one model release may not persist after changes to the model, tools, or memory. Versioned records and replication are necessary to understand such variation.

These limits are part of the proposal's strength. They identify what a future result would mean and where additional evidence is required. A developing field should not use a powerful cultural analogy to blur those boundaries. It should use the analogy to find a question that can be answered more carefully than ordinary impressions of responsiveness allow.

27. A reproducible sequence for the proposed study

PROPOSED RESEARCH — NOT CONDUCTED

PROPOSED STUDY · NOT CONDUCTED

A reproducible evaluation sequence

The first phase would define the world generator, signal alphabet, convention mappings, interaction states, and scoring rules. Independent checks would verify that every trial has the intended information structure. Some trials would be fully specified, some ambiguous, and some noisy. The distinction would be encoded before evaluation. The generator would counterbalance surface forms, object positions, and action frequencies so that simple shortcuts are visible rather than accidentally rewarded.

The second phase would validate the measurement pipeline with baselines. An echo system should detect and repeat forms without solving reference tasks. A lookup system should succeed on fixed familiar mappings. A rule engine with complete state should perform the formal coordination task. Failures in these controls would prompt revision of the materials or scoring. Pilot model outputs could reveal additional ambiguities, but any changes based on them would be documented before the held-out evaluation set is frozen.

Manuscript passage · §27 · wording unchanged

The third phase would compare selected model-mediated arrangements under predeclared conditions. Direct response, structured confirmation, and interactive clarification would be separate interventions. Symbolic inputs would establish a baseline before sensory encodings were added. Sessions would be isolated unless cross-session learning is explicitly tested. All task-relevant inputs, tool calls, and simulator outcomes would be retained in a versioned record. No real external action would be permitted.

The fourth phase would score component and sequence outcomes. Automatic checks would assess signal identity, world constraints, and completed actions. Human review, blinded where practical, would assess explanation fidelity and the specificity of clarification. The analysis would respect world-level clustering and distinguish confirmatory from exploratory comparisons. Missing outputs and disagreements would be reported. The study would not replace uncertainty with a single dramatic percentage.

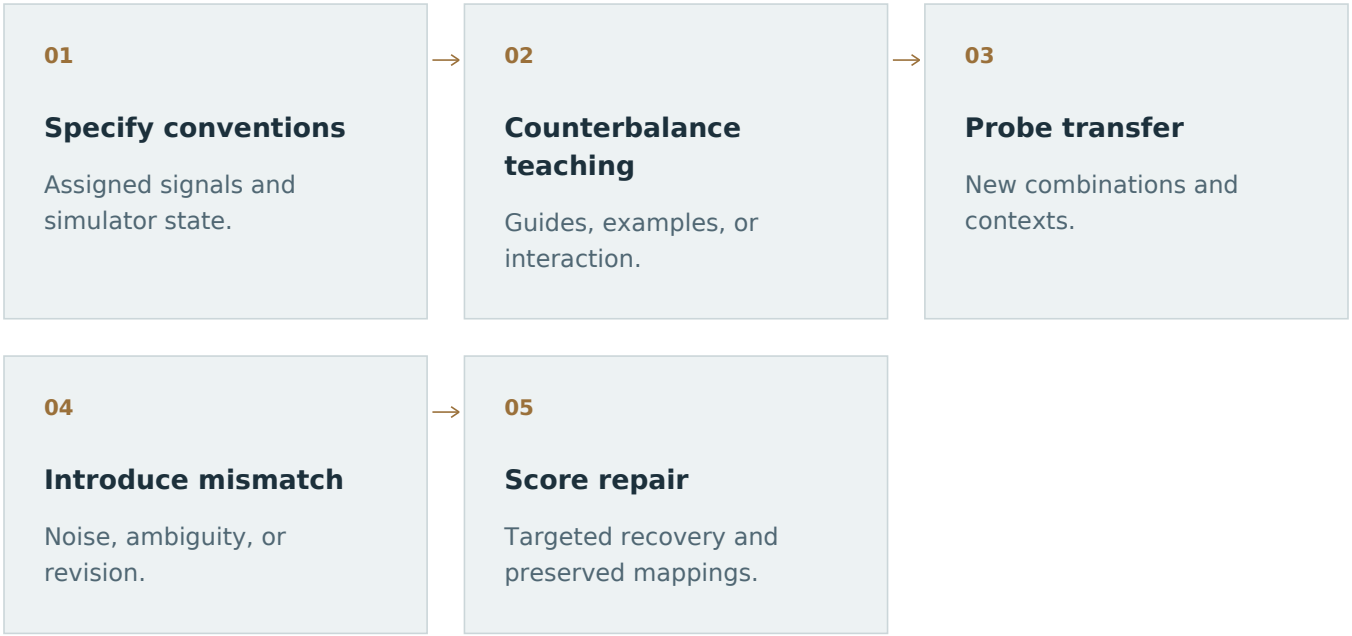
The fifth phase would test repair and convention revision on new combinations. The key question would be whether updates are targeted, retained, and appropriately superseded. A later replication would use newly generated worlds and, where possible, different

implementations. The artifact package would include generators, renderers, protocol definitions, and scoring code so that independent teams could identify remaining shortcuts or reproduce the conditional patterns.

The final report would state what the arrangement demonstrated and what it did not. It might show reliable detection but fragile reference, or strong coordination only with explicit confirmation. Those are useful results. The research would contribute by making the structure of the bridge inspectable, not by declaring that a first reply has resolved the problem of understanding another kind of mind.

FIGURE 2 / CONCEPTUAL SCHEMATIC

The proposed communication evaluation



A study map, not a report of results. Keep channel reliability, semantic adequacy, action validity, and reporting accuracy separate.

Reading aid based on §10, §11, §21, §27. Source paragraphs remain in the full manuscript.

28. Conclusion: the first signal is the beginning of the method

Close Encounters makes communication across unfamiliar forms feel possible. A scientific response should preserve that curiosity while asking what the exchange establishes. Detection, repetition, shared mapping, reference, coordination, and repair are distinct achievements. They can support one another, but none should be silently substituted for all the others. A channel can be perfectly reliable while its users disagree about meaning, and a task can succeed through context or safeguards even when the intended signal was misunderstood.

This paper has proposed a synthetic environment in which those distinctions can be tested. Novel signal assignments, counterbalanced objects, explicit protocol states, and simulated actions make the evidence inspectable. Echo, lookup, rule-based, and model-mediated arrangements provide comparison points. Noise and ambiguity receive different repair options. Convention changes test both flexibility and stability. Provenance traces distinguish the contribution of the model from the contribution of tools, guards, and shared records.

The study remains a proposal. No performance results, human judgments, or claims of validated understanding are reported. Its contribution is the design of a layered evaluation and an argument for matching each measure to its evidential scope. That approach allows quantitative analysis without treating information transfer as a measure of consciousness, and practical coordination without assuming identical internal representations.

Within the Xenopsychology series, this paper addresses the channel through which a bridge can begin. Darmok examines the cultural background that gives expressions their role. Arrival examines the assumptions in the task being asked. Here the question is how signals acquire reference, how commitments become explicit, and how participants discover that an apparently successful exchange has not yet established enough agreement to act.

The shared future of human and artificial systems will depend on more than their ability to answer one another. It will depend on whether they can distinguish receipt from understanding, understanding from authorization, and attempted action from completed action. Those distinctions are less spectacular than the first musical reply in a film. They are also the foundation on which reliable communication can be built. Xenopsychology's task is to make that foundation visible, testable, and open to revision.

“Detection, repetition, shared mapping, reference, coordination, and repair are distinct achievements.”

Rob Emerick · Close Encounters · §28

29. Appendix: a transcript that separates four acknowledgments

A minimal communication audit can use four explicitly different acknowledgment states. A receipt confirms that a signal arrived. An interpretation acknowledgment states which object or action the receiver believes the signal denotes. An authorization acknowledgment confirms that the proposed action is permitted within a specified scope. A completion acknowledgment records that the action occurred in the simulated world. These states should be separately represented even when a user interface normally compresses them into a single affirmative response.

Consider a constructed request concerning the striped triangular object in the virtual tabletop. The receiver first echoes the signal. That supplies evidence of receipt, but it leaves open whether the receiver has associated the signal with shape, texture, position, or the full object identity. An interpretation acknowledgment can name the selected object using an independent identifier. The sender can then compare that identifier with the intended target before the interaction advances toward action. The comparison is meaningful because the synthetic world makes the intended referent inspectable.

Next, suppose the receiver correctly identifies the object but lacks permission to move it to the requested zone. A truthful authorization acknowledgment should state the missing permission rather than treat successful interpretation as permission. If authorization is later granted for a different destination, the update must retain that scope. The system should not generalize permission to every movement of the same object. This creates a clean distinction between understanding the request and being allowed to carry it out.

Finally, a completed move should be recorded by the simulated world's state transition, not merely by the receiver's sentence announcing completion. A tool may reject an attempted move or return an uncertain outcome. The completion message should preserve that information. Otherwise, the communication layer can manufacture apparent success while

the action layer remains unchanged. Such cases are useful negative controls because the signal exchange can sound perfectly coherent despite a failed task.

The transcript format should attach identifiers to the request, interpretation, authorization, attempted action, and observed result. A correction can then specify which element is being repaired. A request to retransmit an unclear signal should not erase a valid authorization record; a revoked authorization should not be treated as evidence that the original signal was misheard. The distinctions support targeted repair rather than a generic restart of the whole conversation.

These acknowledgments are proposed experimental conventions, not a claim that all communication requires this exact protocol. Their value is diagnostic. They reveal how a system can succeed at one level and fail at another, and they give observers a way to locate the failure without inferring shared meaning from mere repetition. A successful study would show where the distinctions improve coordination, where they impose unnecessary burden, and which properties transfer to less structured exchanges.

Open questions

1. Can an evaluation distinguish echoing from conditional use of a convention?

2. What observable outcome demonstrates repair rather than a verbal acknowledgment?

3. Which simple baseline is sufficient for the task being deployed?

References & source scope

The 1977 film’s landing-site communication motif. The article distinguishes channel reliability, task coordination, and reference/repair.

Fiction / work identification

[1] **Steven Spielberg. Close Encounters of the Third Kind (1977). Sony Pictures film page.**

Work identification and official film summary, not a full transcript of the light-and-sound sequence. No dialogue or musical sequence is reproduced.

https://www.sonypictures.com/movies/closeencountersofthethirdkind

Official production commentary

[2] **Amblin. Close Encounters of the Third Kind: official production page.**

Official production context. The new tabletop communication protocol is original conceptual work, not a reconstruction of the film’s technology.

https://amblin.com/movie/close-encounters-of-the-third-kind/

Research

[3] **Shannon (1948). A Mathematical Theory of Communication. Bell System Technical Journal 27.**

Primary paper, reprint hosted by Harvard. Channel and information concepts are distinguished from a theory of shared semantic interpretation.

https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf

[4] **Clark & Brennan (1991). Grounding in Communication. In Perspectives on Socially Shared Cognition.**

Primary chapter hosted by Stanford. Task-relative grounding and the role of communication media provide intellectual lineage; our artificial-world protocols are separate proposals.

https://web.stanford.edu/~clark/1990s/Clark,%20H.H.%20_%20Brennan,%20S.E.%20_Grounding%20in%20communication_%201991.pdf

[5] **Dingemanse et al. (2015). Universal Principles in the Repair of Communication Problems. PLOS ONE 10(9): e0136100.**

Research on human conversational repair. Cited as a methodological precedent; it does not establish that identical principles or outcomes hold for AI.

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0136100

[6] **Bender & Koller (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. ACL, 5185–5198.**

Authors' abstract and bibliographic record. A position argument about form and meaning; not treated as a theorem settling every multimodal or interactive system. DOI: 10.18653/v1/2020.acl-main.463.

https://aclanthology.org/2020.acl-main.463/

Publication record

Rob Emerick (2026). Close Encounters — Finding a Shared Cognitive Channel. XENO-WP-2026-004, conceptual manuscript R3; scholarly reading edition R4, author attribution R5. Xenopsychology. Not peer reviewed. Reading edition prepared 2026-09-17.

AI-assisted drafting and editorial preparation. Human scholarly review remains required. The series identifier is internal, not a DOI. Reading aids are editorial syntheses of the manuscript, not new findings. Original main-text paragraphs, abstract, open questions, and source records are preserved.

Abstract: https://xenopsychology.com/insights/close-encounters-shared-cognitive-channel

Full paper: https://xenopsychology.com/insights/close-encounters-shared-cognitive-channel/paper

Author note

Rob Emerick

Co-founder, Xenopsychology · 2026–present · Systems architect

ORCID iD: https://orcid.org/0009-0006-1269-4216

Rob Emerick is a systems architect and co-founder of Xenopsychology whose work spans artificial cognition, data integration, and animal-welfare infrastructure. He founded Planet IDX and was the sole creator of REML (Real Estate Modular Language), an interpreted language developed to integrate disparate real-estate listing systems. He also founded Pantheon Golem, where he develops AI systems informed by structured analysis of fictional characters and worlds. Through Rescue Nexus and Chipped Pets, he is developing animal-rescue infrastructure, a shared ontology for shelter data integration, and pet-identification technology. His broader work includes veterinary-forensics software and veterinary hematology technology under development.

About this attribution

The byline and original manuscript pull quotes are attributed to Rob Emerick at his request. Quotations and references from outside sources retain their original attribution. The biography is interview-based; no degrees, appointments, contribution roles, or independent validation are inferred.

AI-assisted drafting and editorial preparation. Human scholarly review remains required. This remains a conceptual working paper, not a peer-reviewed publication or a report of completed experiments.

Biography: https://xenopsychology.com/people/rob-emerick