

The Reality and Recognition of Suffering: A Generative Relational Epistemology

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Suffering plays a central role in moral judgment, humanitarian practice, animal ethics, and emerging debates concerning artificial agents. Yet suffering is rarely available to an observer as a directly accessible fact. What is ordinarily recognized as suffering is mediated through bodily expression, language, behavioral response, functional disruption, prior knowledge, and historically developed interpretive categories. These difficulties become especially visible when the subject of recognition differs substantially from the observer in embodiment, communication, cognition, or mode of existence. This paper develops a preliminary Generative Relational account of the recognition of suffering. Rather than resolving the ontological status of phenomenal experience, the analysis suspends that question and distinguishes several analytically separable dimensions: phenomenal suffering, manifested suffering, functional adverse states, and relational or trajectory-level harm. It then examines how an observer moves from manifestations to recognized suffering through an interpretive system, and how such recognition can fail when familiar human expressions are treated as privileged indicators. Cross-species, artificial-agent, and hypothetical alien cases are used to expose two related problems: human-readable suffering bias and interpretive inversion, in which manifestations of vulnerability may be classified as threat, aggression, or dysfunction. The analysis further considers conditions under which relational dynamics may remain partially observable even when semantic interpretation is unavailable. Persistent avoidance, costly attempts to terminate a condition, changes in future capacities, and recurrent relational effects may provide evidence about adverse states without providing transparent access to phenomenal experience. Such evidence, however, does not collapse the distinction between relational reconstruction and first-person suffering. To compare heterogeneous cases, the paper therefore treats similarity as dependent on specified relational distinctions rather than complete structural identity, raising the question of what is preserved when different manifestations are treated as epistemically comparable. The paper concludes by presenting suffering recognition as historically and interpretively situated rather than as a direct reading of an inner state. Its contribution is epistemic rather than prescriptive: it does not infer duties of aid or permissions to intervene from recognized suffering. Instead, it clarifies the conditions, limits, and possible failures through which suffering becomes recognizable across heterogeneous relational systems.

Keywords: generative relational epistemology; suffering; recognition; relational reality; manifestation; phenomenal experience; interpretive systems; cross-species recognition; artificial agents; semantic opacity; interpretive inversion; heterogeneous subjects

1 Suffering as an Epistemic Problem

This section establishes the epistemic object of the paper. Its purpose is not to define suffering exhaustively, to settle the ontology of phenomenal experience, or to derive moral duties from the presence of suffering. It instead isolates a logically prior problem: suffering often enters moral judgment, humanitarian practice, medicine, animal welfare, and emerging discussions of artificial agents as though it were already available as an identified fact, while the production of that identification is itself mediated by observation, testimony, interpretation, background knowledge, and historically developed concepts. The discussion proceeds in three stages. It first locates suffering recognition within several practical domains in which judgments about another's adverse condition matter. It then distinguishes this practical problem from radical scepticism about other minds and focuses on the more limited question of how claims about another's experience become warranted and corrigible. It finally specifies the narrower Generative Relational inquiry pursued in the remainder of the paper.

1.1 Suffering Recognition across Normative and Practical Domains

Suffering has an unusual epistemic position in practical reasoning. It frequently functions as a premise before the conditions under which that premise was recognized are made explicit. Humanitarian practice offers a clear example. The International Red Cross and Red Crescent Movement formulates humanity in terms of preventing and alleviating human suffering and protecting life and health, while impartiality directs attention to need and the urgency of distress [23]. Whatever normative justification is ultimately given for these commitments, their operation presupposes some capacity to identify conditions as suffering, distress, injury, need, or vulnerability. A humanitarian actor who cannot distinguish injury from ordinary variation, distress from threat, or incapacity from refusal cannot simply proceed from a principle of humanity to an appropriate response. The normative principle therefore depends, in practice, on an epistemic layer that is conceptually distinguishable from the principle itself.

Clinical pain assessment makes this dependence especially visible. The revised International Association for the Study of Pain definition describes pain as a sensory and emotional experience and emphasizes that pain is personal, that pain and nociception are distinct, and that pain cannot be inferred solely from neural activity. Its accompanying notes also state that verbal description is only one among several pain behaviors and that inability to communicate does not negate the possibility of pain in a human or nonhuman animal [35]. These points do not provide a theory of suffering in the broader sense used in this paper. Pain and suffering are not treated here as synonyms. They do, however, show in a mature field of assessment why a morally and clinically important adverse state cannot be reduced either to one outward sign or to one measurable physiological process. Recognition requires relations among experience, manifestation, context, prior knowledge, and an observer's interpretive practices.

Research on pain communication develops this point further by treating pain expression and pain recognition as socially situated processes rather than as a simple transmission of an internal magnitude. Hadjistavropoulos and colleagues formulate pain communication in terms of interactions among biological, psychological, social, expressive, and interpretive processes affecting both persons in pain and those who respond to them [20]. This literature is useful here not because a model developed for human pain can be exported unchanged to animals, artificial agents, or hypothetical extraterrestrial subjects, but because it makes a structural point that survives the domain change: the state of one subject and the recognition achieved by another are not identical events.

For the present inquiry, it is therefore useful to reserve a notation for recognized suffering that does not prejudge the underlying phenomenal reality. Let O_B denote some set of observable manifestations or traces associated with a subject B , and let I_H denote a human interpretive system operating with available evidence, concepts, and background knowledge. A minimal schematic relation is

$$O_B \xrightarrow{I_H} \hat{S}_B.$$

Here $\hat{S}_B$ denotes a recognition or attribution of suffering by the interpreter; the hat is important. The expression does not assert that $\hat{S}_B$ is identical to the total state of B , nor that the interpreter has transparent access to a first-person phenomenal condition. The same notation can accommodate a correct recognition, an under-recognition, an over-recognition, or a classification failure. It can also be extended later to interpreters other than humans.

This distinction matters because practical reasoning often compresses the epistemic transition. A sequence such as “ B is suffering; therefore B requires protection” can conceal at least two different questions: first, what warrants the claim that B is suffering; second, what normative consequences follow if that claim is warranted. The present paper addresses the first problem. It intentionally leaves the second to subsequent work. This separation does not imply that epistemology and normativity are independent in social life. Interpretive practices are themselves shaped by institutions, norms, expectations, and prior relations. The analytical separation is methodological: before evaluating the normative bridge from suffering to aid, it is necessary to examine how the suffering premise becomes available for use.

The same issue appears even where no formal humanitarian institution is involved. In medicine, caregivers must assess subjects whose reports may be incomplete or unavailable. In animal welfare, observers infer adverse conditions without linguistic testimony in a shared human language. In discussions of artificial agents, behavioral outputs, self-reports, architectural properties, and persistent relational patterns may be proposed as evidence while the relevance of each remains contested. Hypothetical alien cases radicalize the same structure by removing shared embodiment, evolutionary history, language, and familiar expressive conventions. These cases differ profoundly in their empirical basis, and this paper will not treat them as interchangeable. Their comparative value lies in exposing which inferential supports are present in familiar human cases and which disappear as interpretive distance increases.

1.2 Access to Another’s Experience

The epistemic difficulty identified here belongs to the broad territory traditionally associated with knowledge of other minds, but the paper does not adopt radical scepticism as its starting point. Philosophical discussions of other minds distinguish questions about whether knowledge of another mind is possible from questions about the sources, grounds, and possible errors involved in such knowledge. Avramides characterizes the familiar epistemological problem in terms of how beliefs about another’s thoughts and feelings can be justified, noting the roles assigned in the literature to analogy, inference to the best explanation, testimony, criteria, and perceptual accounts [3]. The present inquiry is closest to the less radical source-and-error problem. Ordinary life, medicine, and humanitarian practice cannot suspend action until solipsism is refuted; they require fallible practices for recognizing another’s condition.

A useful asymmetry nevertheless remains. A subject may have first-person access to an experience in a manner unavailable to an observer, whereas the observer ordinarily

encounters testimony, bodily expression, behavior, physiological measurements, environmental circumstances, and the history of the interaction. Even this contrast should not be overstated: first-person interpretation can be uncertain, socially mediated, or conceptually incomplete, and testimony can itself be an important source of knowledge rather than merely a symptom from which an inner state is inferred. The point required here is narrower. Another subject's experience is not simply transferred into the observer unchanged. Recognition occurs through epistemic relations whose reliability may vary.

The revised IASP account is instructive precisely because it refuses two opposite reductions. On one side, pain is not identified with nociceptive activity; on the other, verbal report is not made the only admissible route to pain recognition [35]. This creates a practical space in which testimony can carry distinctive authority while nonverbal evidence remains relevant. The structure is important beyond clinical pain even though the substantive criteria cannot simply be generalized. If an observer treats only human-readable verbal or expressive behavior as evidence, genuine adverse states may become invisible when subjects cannot use those channels. If, conversely, any behavior that resembles a familiar pain expression is treated as sufficient proof of a homologous experience, resemblance can be overinterpreted. Recognition therefore faces both false-negative and false-positive risks.

The communication literature similarly shows that expression and decoding are shaped by context. Pain behavior can be encoded differently across subjects, while observers interpret behavior through expectations, learned categories, roles, and situational information [20]. A recognition claim is thus neither a photograph of an inner object nor a free construction unconstrained by the subject. It is better treated, at this preliminary stage, as an epistemic achievement produced within a relation between manifestations, an interpreter, and a body of contextual evidence. Some achievements will be reliable; others will fail. The task is to make the structure of that fallibility analyzable.

This framing also prevents radical heterogeneity from being treated as though it created an entirely new problem. Human-human recognition already contains mediation. Shared embodiment, language, medical knowledge, and social conventions make many recognitional inferences comparatively well supported, but they do not eliminate the difference between another's experience and an observer's recognition of it. Animal, artificial, and alien cases progressively remove some of these shared supports. They therefore function as stress tests of an epistemic architecture that is already present in ordinary human cases. The relevant contrast is not between perfectly transparent human recognition and completely opaque nonhuman recognition. It is between relational settings with different degrees and kinds of evidential and interpretive support.

The distinction between pain and suffering is also important at this stage. Pain is a relatively developed object of clinical and scientific inquiry, whereas suffering may include forms of distress, loss, fear, deprivation, disruption, or adverse transformation that are not exhausted by pain. Accordingly, the paper will not infer a general theory of suffering directly from pain science. Instead, pain research supplies disciplined examples of a broader epistemic structure: a personally significant adverse state may have manifestations and correlates, but neither a single manifestation nor a single correlate exhausts the state to be recognized. Later sections will use this structure to distinguish phenomenal, manifestational, functional, and relational dimensions without assuming in advance that all four coincide.

1.3 Epistemic Scope of the Present Inquiry

The narrower research question can now be stated. Given an observer or interpreter *I*, a subject *B*, a historically situated body of evidence, and a relational context, under what conditions can a claim that *B* is suffering become warranted, comparable, revisable, or contestable without being treated as transparent access to *B*'s phenomenal life? The question is deliberately framed around recognition rather than detection. "Detection" can suggest that suffering is already available as a fixed observable variable awaiting measurement. "Recognition" leaves open the possibility that the relevant categories, evidential relations, and comparison standards themselves require interpretation and may change historically.

The Generative Relational approach contributes three methodological commitments to this inquiry. First, it suspends ontological settlement where settlement is not required for the epistemic question. The paper need not decide, at the outset, a complete metaphysics of consciousness in order to distinguish what is observed from what is inferred. Second, it treats recognition as relationally situated. Manifestations acquire evidential significance within histories of interaction, interpretive practices, bodies of knowledge, and structures of comparison. Third, it treats these structures as potentially generative and revisable. New evidence can matter, but so can new concepts, new distinctions, and new ways of grouping previously incomparable cases.

These commitments impose limits as well as possibilities. Relational analysis must not be used to replace phenomenal suffering with functional disruption. An artificial system that persistently avoids a state, an animal that changes behavior after injury, or an alien that undergoes a dramatic relational transformation does not thereby become proven to have a human-like phenomenal experience. Conversely, lack of a familiar manifestation does not establish absence of an adverse experience. The purpose of the framework is to preserve these distinctions long enough to study the evidential relations among them.

The remainder of the paper develops this inquiry in stages. The next section introduces the distinction among ontological, relational, and manifestational reality. The following section separates phenomenal suffering, manifested suffering, functional adverse states, and relational or trajectory-level harm. Subsequent sections begin with human-human recognition and then progressively examine cross-species and radically heterogeneous cases, including interpretive inversion and semantic opacity. The paper then turns to cross-interpretive comparison, where similarity will be formulated relative to specified preserved structures rather than assumed as complete identity, and finally to the historical development of recognizability. Throughout, thought experiments involving artificial or alien subjects function as conceptual stress tests; empirical claims about animals, clinical populations, or current artificial systems require independent evidence.

The resulting scope is intentionally narrower than a complete ethics of suffering. The paper does not ask whether every recognized sufferer has the same moral status, whether suffering is sufficient to generate an obligation of aid, whether a third party has authority to intervene, or how conflicting claims should be balanced. Those problems become visible only after the epistemic premise has been made explicit. The present task is prior: to analyze how a proposition of the form "*B* suffers" becomes available to a situated interpreter and what can be said about the reliability, limits, and revisability of that recognition. In this sense, suffering is treated first as an epistemic problem before it is used as a normative premise.

2 Three Layers of Reality in GR

The epistemic problem developed in Section 1 requires a distinction among several senses in which something may be called real. Without such a distinction, claims about suffering can shift unnoticed between a phenomenal condition, the causal and relational effects associated with that condition, and the manifestations through which an observer recognizes it. The Generative Relational framework uses a three-layer working distinction to prevent this collapse. It leaves ontological reality unsettled where the inquiry does not require a metaphysical verdict, treats relational reality as the level at which entities, states, relations, and classifications can make sustained differences within a generative system, and treats manifestational or representational reality as the level at which those differences become observable, expressible, encoded, or otherwise available to an interpreter. These layers are analytically distinct but not isolated. Manifestations occur within relational systems, relational structures condition what can be manifested and recognized, and ontological claims may later be proposed about either. The present paper, however, does not require those claims to be settled in advance.

2.1 Ontological Suspension

The first commitment is methodological rather than metaphysical. When the paper writes

$$O(B) = ?,$$

where $O(B)$ stands schematically for the ultimate ontological status of a subject or state, the question mark is not an assertion that no answer exists. It marks a suspension of judgment where the answer is not required for the analysis being performed. In the present context, the paper does not need to decide a complete metaphysics of consciousness, phenomenal subjectivity, biological life, artificial mentality, or extraterrestrial experience before it can distinguish an observed manifestation from an inferred adverse state. Ontological suspension therefore functions as a constraint on inference: an epistemic claim should not obtain more metaphysical content than the evidence and argumentative task require.

This position should be separated from both ontological anti-realism and ontological reduction. To suspend the ontology of phenomenal suffering is not to say that phenomenal suffering is unreal, socially constructed, or reducible to observable behavior. Nor is it to say that relations are the only things that ultimately exist. The distinction is particularly important because nearby positions in philosophy of science sometimes combine epistemic restraint with stronger claims about structure. Epistemic structural realism, for example, has been formulated as a position on which one may commit to structural or relational content while remaining agnostic about the intrinsic nature of unobservable entities [45]. The present use of suspension is weaker. GR does not infer from limited access to intrinsic nature that structure alone exhausts reality, and it does not require the thesis that relations are ontologically prior to relata. Structural realism is therefore a useful neighboring comparison, not the metaphysical foundation of the present account.

The same caution appears in earlier work on Generative Relational Epistemology, where the claim that a concept exceeds any single symbolic representation was explicitly separated from a verdict about whether that concept exists independently of its representations, emerges from relations among them, or should have its independent status left unsettled [22]. The present paper adopts the same discipline. A subject's apparent distress may exceed any one manifestation available to an observer, but this epistemic excess does not by itself establish what the underlying phenomenal state ultimately is.

The distinction allows the analysis to remain informative without converting incomplete access into either metaphysical realism or metaphysical denial.

Ontological suspension also has a practical epistemic consequence. If the analysis began by stipulating that only organisms with a known biological substrate can suffer, artificial and alien cases would be excluded by definition before their evidential structure was examined. If it began by stipulating that any sufficiently complex or persistent adverse dynamics instantiate suffering, the paper would have solved the hard question by terminological expansion. Both moves prejudge what the inquiry is meant to clarify. Suspension instead permits heterogeneous cases to be compared at the levels for which evidence is available while preserving the distinction between those comparisons and claims about phenomenal identity.

The methodological form can be expressed as

$$\text{epistemic analysis of } B \not\Rightarrow \text{ontological settlement of } B.$$

A later theory may propose stronger metaphysical commitments. The present framework remains compatible with several such possibilities provided they preserve the distinctions required for recognition. The immediate task is therefore not to decide what suffering is in itself, but to identify what can be responsibly said when different layers of evidence, effect, and manifestation are available.

2.2 Relational Reality

The second layer concerns what is real in virtue of its operation within a generative relational system. The working idea is that a state, distinction, relation, institution, representation, or attributed status may be relationally real when it persistently makes a difference to the system's interactions, constraints, trajectories, or future possibilities. This does not require the item to possess an independently settled ontology. Legal obligations, money, institutional roles, reputational classifications, and social identities provide familiar examples: their metaphysical status is contested in different traditions, yet changes in them can alter access, expectation, action, resource flow, and future relations. In that limited sense, their relational efficacy is not exhausted by whether an observer currently endorses a particular metaphysical account of them.

For the purposes of suffering recognition, relational reality is important because adverse conditions can have persistent effects even when their phenomenal character is inaccessible. Suppose that a subject B undergoes a state $x_B(t)$ after which its communication patterns, available actions, resource relations, memory continuity, or future trajectories are systematically altered. The paper may describe these changes without treating them as transparent measures of first-person suffering. Schematically,

$$x_B(t) \longrightarrow \Delta R_B(t : t + \tau),$$

where $\Delta R_B(t : t + \tau)$ denotes a change in the relational trajectory associated with B over a later interval. The expression records an effect structure. It does not assert

$$\Delta R_B(t : t + \tau) = S_B^{\text{phen}}.$$

The distinction between those statements is central to the paper.

Relational reality is therefore neither a substitute name for phenomenal reality nor a lower-resolution version of it. A relationally real effect may occur without suffering; suffering may occur without some effect an external observer can readily detect; and the same phenomenal condition may enter different relational trajectories in different contexts. The analytical value of the relational layer lies precisely in allowing the paper

to say more than “nothing can be known without first-person access” while saying less than “observable adverse dynamics prove phenomenal suffering.” Later sections will use this middle position to examine avoidance, costly departure, persistent disruption, and capability loss as evidence about adverse states whose phenomenal interpretation remains open.

The term “relational” also marks a departure from a purely entity-centered analysis. Recognition often depends not only on properties attributed to B but on relations among B , its environment, other subjects, prior states, available alternatives, and the interpreter. A vocalization may be ordinary in one relational context and evidence of severe distress in another; immobility may indicate rest, defensive freezing, incapacity, or a species-specific adaptive response; repeated attempts to leave a state may acquire evidential significance only when the relevant costs and alternatives are known. The object of inquiry is therefore not an isolated sign but a structured history of relations in which signs and effects acquire significance.

This generative emphasis also introduces time. A relationally real pattern need not be reducible to a single state at a single moment. Some conditions become epistemically legible only across trajectories. An isolated action can be ambiguous, while a persistent sequence of avoidance, failed recovery, altered interaction, and long-horizon capacity loss can support a stronger inference about an adverse condition. Conversely, a dramatic instantaneous manifestation can be misleading when detached from the longer dynamics that generated it. The relational layer thus prepares the later distinction between state-level observations and trajectory-level harm.

The framework should not be read as claiming that every causally efficacious pattern deserves the name “suffering.” Many relationally real structures are ethically or phenomenally neutral. Nor does relational efficacy alone determine moral standing. The present claim is only epistemic: where direct phenomenal access is unavailable, persistent relational structure can constitute part of the evidence from which an observer reconstructs an adverse condition. What that evidence warrants must be assessed case by case, and the relevant equivalence criteria become especially important when heterogeneous subjects are compared.

2.3 Manifestational and Representational Reality

The third layer concerns what becomes available to an interpreter through manifestation or representation. A manifestation can be a vocalization, gesture, bodily change, behavioral pattern, self-report, physiological measurement, machine output, log entry, environmental trace, image, or other observable event. A representation can be linguistic, mathematical, graphical, diagnostic, computational, institutional, or classificatory. Scientific practice itself relies on heterogeneous representational forms, and philosophical work on scientific representation emphasizes that representations are not identical to their targets: they are used by agents for purposes within contexts, can be more or less accurate, and can misrepresent while remaining representations [16]. This point is directly relevant to suffering recognition.

Let M_B denote a manifestation associated with subject B . Recognition can then be written more precisely as

$$M_B \xrightarrow{I_H} \hat{S}_B.$$

The arrow denotes an interpretive transition, not an identity. A cry can be a genuine acoustic event even if an observer misclassifies its significance. A change in body color can be a genuine physiological manifestation even if it is interpreted as aggression rather than distress. An artificial agent’s sentence “I am in pain” can be a genuine linguistic output even if its relation to phenomenal experience remains unresolved. In each case,

the manifestation is real as an event in the relational system while the representational or evidential relation between that event and suffering remains open to evaluation.

This distinction prevents two opposite mistakes. The first treats manifestation as transparent revelation. On that view,

$$M_B \Rightarrow S_B$$

whenever the manifestation resembles a familiar indicator of suffering. Such a rule invites anthropomorphic over-attribution and becomes especially fragile across species or artificial architectures. The second mistake treats mediated access as epistemically worthless. On that view, because M_B is not identical to phenomenal suffering, no justified inference can be drawn from manifestation at all. Ordinary medicine, communication, and animal-welfare practice would become impossible under such a standard. The paper instead treats manifestations as defeasible evidence whose significance depends on the interpreter, surrounding relations, comparative knowledge, and historical context.

The interpreter is therefore part of the epistemic configuration. We can write

$$\hat{S}_{B,H,t} = I_{H,t}(M_B, E_t, C_t),$$

where E_t denotes the evidence available at time t and C_t denotes the conceptual resources through which that evidence is classified. The notation is schematic rather than a quantitative model. Its purpose is to make visible that recognition can change even when the past manifestation itself does not. New evidence can alter its meaning; new concepts can create distinctions previously unavailable; and a later interpreter may recognize that an earlier category combined states that should have been separated. The manifestation remains part of the historical record, while the recognized suffering attribution is temporally and interpretively indexed.

This temporal indexing matters for the paper's later discussion of recognizability. Consider a nonhuman subject whose severe adverse state is expressed primarily through a color change that humans initially interpret as a neutral physiological variation. At t_0 , the manifestation may be fully observable while its significance remains unavailable. At t_1 , comparative observation, intervention, or communication may reveal that the color change reliably accompanies costly avoidance, functional disruption, or other adverse dynamics. The relational and manifestational histories have not been created retroactively; what changes is the interpretive system capable of recognizing their relation. In this sense, ethical or clinical visibility can lag behind the events to which it refers.

The representational layer also allows testimony to be handled without either granting it infallibility or reducing it to mere behavior. A subject's report can be evidentially distinctive because it is a representation produced from the subject's own standpoint. Yet reports can be constrained, deceptive, confused, culturally patterned, architecturally generated, or conceptually limited. The present paper therefore does not rank testimony and nonverbal manifestation by a single universal hierarchy. It asks instead what relation each source bears to the state being recognized and what reasons justify treating that relation as informative in the case at hand.

The three-layer distinction can now be summarized without collapsing the layers. Ontological reality concerns what the subject or state ultimately is; this paper suspends that question where unnecessary. Relational reality concerns the sustained differences that states, relations, and classifications make within a generative system. Manifestational or representational reality concerns the events and encodings through which an interpreter gains epistemic access to those differences. The layers interact, but no arrow among them is automatic:

$$\text{manifestation} \neq \text{relational effect} \neq \text{phenomenal state}.$$

The remainder of the paper builds on this separation. The next section will distinguish dimensions of suffering themselves, while later sections examine the evidential relations through which manifestations and relational trajectories are used to recognize them.

3 Dimensions of Suffering

The three-layer distinction developed in Section 2 separates ontological, relational, and manifestational questions, but it does not yet specify what kinds of phenomena are being distinguished when observers speak of suffering. A further analytical separation is therefore required. In ordinary language, the word “suffering” can refer to a lived negative experience, its outward expression, a loss of function, a prolonged disruption of a life, or some combination of these. Clinical and philosophical literature has long warned against treating pain, distress, bodily dysfunction, and suffering as interchangeable. The revised International Association for the Study of Pain definition, for example, treats pain as a personal sensory and emotional experience, distinguishes pain from nociception, notes that pain can adversely affect function and social and psychological well-being, and explicitly rejects verbal description as a necessary condition of pain expression [35]. Cassell likewise distinguished suffering from physical distress and emphasized that suffering may concern threats to the integrity of a person rather than bodily pain alone [8]. These traditions do not provide a single theory that the present paper adopts. They nevertheless support the need to disaggregate phenomena that are often compressed into one term.

The present analysis uses four dimensions: phenomenal suffering, manifested suffering, functional adverse states, and relational or trajectory-level harm. The dimensions are not proposed as four mutually exclusive kinds of entity, nor as an exhaustive taxonomy of every phenomenon that may be called suffering. They are analytical coordinates designed to prevent evidential movement from being mistaken for identity. A single episode may involve all four dimensions, while another case may provide evidence for only one or two. The central discipline is to specify which dimension is being observed, which is being inferred, and which inferential bridge remains open.

Table 1 is deliberately non-hierarchical. It does not imply that one dimension is more real than another, or that observation should proceed from the fourth row toward the first as if reconstructing a hidden scalar quantity. The dimensions answer different questions. A phenomenal question concerns what an adverse state is like for the subject. A manifestational question concerns how that state becomes available to an observer. A functional question concerns what capacities or regulatory processes are disrupted. A relational question concerns how the condition changes the subject’s position, interaction, continuity, and future possibilities over time. Their evidential relations must be argued rather than assumed.

3.1 Phenomenal Experience as the First-Person Dimension

Phenomenal suffering refers here to the adverse character of experience from the subject’s own standpoint. The term is intentionally broader than nociception and narrower than every condition that an observer might judge harmful. In human pain research, the distinction between nociceptive activity and pain provides a useful analogue: the revised IASP definition characterizes pain as an unpleasant sensory and emotional experience and explicitly states that pain cannot be inferred solely from activity in sensory neurons [35]. The point for the present paper is methodological. Even where there are

Table 1: Working dimensions used in the analysis of suffering.

Dimension	Primary object of analysis	Typical mode of access	Analytical caution
Phenomenal	First-person unpleasant or adverse experience	Subjective report where available; indirect inference otherwise	Not reducible to no-ciception, behavior, or function
Manifestational	Expression or representation associated with an adverse state	Language, gesture, facial or bodily change, signal, output, trace	Expression can be absent, ambiguous, strategic, or misread
Functional	Disruption of capacities, integrity, regulation, or performance	Behavioral, physiological, clinical, computational, or ecological measures	Functional loss need not imply phenomenal suffering
Relational / trajectory-level	Persistent alteration of relations, options, participation, continuity, or future development	Longitudinal and contextual reconstruction across interactions	A harmful trajectory is not identical to a first-person experience

measurable causal processes associated with an adverse condition, the first-person character of that condition is not identical to those measurements.

The use of “phenomenal” does not commit the paper to a complete theory of consciousness. It marks the epistemic location of the claim. If subject B undergoes a first-person adverse experience at time t , this may be represented schematically as

$$S_B^{\text{phen}}(t).$$

The notation does not assign a numerical magnitude, and it does not imply that experiences are directly commensurable across subjects. It records only that the claim concerns what is experienced by B , rather than what is externally observed about B . For another observer H , the relevant epistemic problem is that

$$S_B^{\text{phen}}(t)$$

is not ordinarily available as an unmediated public object. H encounters reports, expressions, measurements, histories, and effects from which an attribution is formed.

This first-person dimension is one reason why pain and suffering resist simple behavioral definitions. A subject may experience severe pain while attempting to suppress visible expression. A subject may also display behavior that resembles pain for reasons other than pain. The communication literature models this separation explicitly by distinguishing an internal experience, its encoding into expressive behavior, and an observer’s decoding of that behavior [19, 20]. The possibility of distortion can enter at each transition. A lived state can be only partly expressed; expression can be shaped by social expectations or strategic reasons; and an observer can misread an otherwise informative manifestation.

Suffering also exceeds pain. Cassell’s influential clinical account treated suffering as a phenomenon concerning persons whose perceived integrity is threatened, and explicitly separated suffering from physical distress alone [8]. Later philosophical and bioethical work has criticized aspects of that account, including the difficulty of defining the distress on which it relies [44]. The disagreement is useful for the present purpose because

it shows that even within human medicine there is no simple identity among bodily damage, pain, distress, and suffering. The paper therefore does not define phenomenal suffering by tissue injury, pain intensity, or functional incapacity. It treats these as possible correlates, causes, consequences, or evidential sources whose relations may vary by case.

This restraint becomes essential in heterogeneous cases. If an artificial agent persistently reports an adverse internal condition, or an alien organism displays a state that humans later learn to treat as severe distress, the present framework does not infer phenomenal suffering merely from the existence of the report or pattern. At the same time, it does not define phenomenal suffering out of consideration by demanding a specifically human substrate or human form of expression. Ontological suspension from Section 2.1 remains in force. The first-person dimension is retained as a distinct target of inquiry even when the evidence available for it is incomplete.

The analytical consequence is that evidence may support different strengths of claim. In some human contexts, a competent first-person report may provide strong grounds for attributing pain or suffering. In other contexts, communication may be impossible or its semantics uncertain. The evidential basis then shifts toward manifestations, functional change, or relational history. Such a shift does not transform those other dimensions into phenomenal experience. It changes the type of evidence from which a phenomenal attribution might later be considered.

3.2 Manifestation as the Observable Dimension

Manifested suffering concerns the observable or representational forms through which an adverse state becomes available to an interpreter. These forms include self-report, vocalization, facial expression, posture, movement, withdrawal, physiological display, symbolic communication, machine-generated language, logs, and other signals. The category is intentionally broad because the epistemic problem is not restricted to one communication channel. The same underlying adverse state, if there is one, can be manifested differently across individuals, cultures, species, developmental stages, architectures, and situations.

The communication model of pain provides a useful structure for this dimension. Pain-related experience is not directly transferred from one subject to another; it is encoded through expressive behavior and then decoded by observers within social and contextual conditions [19, 20]. In the notation introduced earlier,

$$S_B^{\text{phen}} \longrightarrow M_B \xrightarrow{I_H} \hat{S}_{B,H},$$

where M_B denotes a manifestation and $\hat{S}_{B,H}$ denotes the suffering attribution formed by observer H . Neither arrow is guaranteed to preserve all relevant information. The first can be weak because a subject does not, cannot, or chooses not to express a state. The second can be weak because the observer lacks the conceptual, contextual, or species-specific resources necessary to interpret the manifestation.

The revised IASP notes are particularly important here because they explicitly state that verbal description is only one of several behaviors through which pain may be expressed and that inability to communicate does not negate the possibility of pain in a human or nonhuman animal [35]. This prevents verbal fluency from functioning as a hidden criterion of recognizability. It also illustrates a broader point: the absence of a familiar manifestation cannot by itself establish the absence of the target condition.

Animal pain assessment makes the interpretive character of manifestation especially visible. Because nonhuman animals cannot provide the same kind of linguistic self-report available in many adult human contexts, assessment relies on combinations of

behavioral and physiological indicators. Facial-expression or grimace scales have been developed for several species, but their reliability and diagnostic accuracy vary, and systematic review literature notes limitations in reference standards and validation [13]. The important point is not that facial expression is unreliable; it is that a manifestation acquires evidential force within a learned interpretive infrastructure. Observers must discover which action units, postures, vocalizations, or behavioral changes covary with relevant adverse states in a given species and context.

This structure prevents “human-readable suffering” from being treated as a neutral epistemic baseline. A manifestation that resembles human crying can be immediately salient to a human observer because the decoding infrastructure is already rich. A color change, chemical signal, change in electric field, or machine-state output may initially be opaque, not because it carries less information about the subject, but because the observer lacks the corresponding interpretive mapping. The distinction will become central in the later sections on cross-species recognition and interpretive inversion, where the paper examines cases in which unfamiliar manifestations are misread through human categories.

Manifestational evidence can also be intentionally or structurally shaped. A subject may exaggerate a signal, suppress it, produce it under instruction, or lack control over it. Artificial systems sharpen this issue because an output can be generated by architecture, prompting, training, strategic planning, or internal states that humans do not know how to interpret. The sentence “I am suffering” is therefore a manifestation with potentially important evidential status, but the sentence is not itself the phenomenal state whose existence it represents. Conversely, the inability to produce such a sentence is not evidence of phenomenal absence.

The epistemic task is therefore not to rank all manifestations on one universal ladder but to reconstruct their relation to the state under investigation. This requires context, history, comparative knowledge, and, where possible, converging evidence from other dimensions. Manifested suffering is indispensable to recognition because it is one of the principal interfaces through which suffering becomes socially visible. Its importance lies in this mediating role, not in an identity between expression and experience.

3.3 Functional Adversity as Systemic Impairment

A third dimension concerns functional adverse states: disruptions in the capacities, integrity, regulation, or performance of a subject or system. In human medicine these may include impaired mobility, sleep disruption, reduced ability to work or communicate, cognitive impairment, loss of self-care capacity, or deterioration in social participation. In animals they may include altered feeding, locomotion, grooming, play, or other species-relevant behavior. In artificial systems analogous descriptions might concern loss of memory continuity, communication, processing capacity, error regulation, or access to functions necessary for continued operation. The analogies are functional rather than ontological; a reduction in computational performance is not thereby the same phenomenon as human pain.

Functional adversity is analytically useful because it is often more publicly measurable than first-person experience. It can also persist across time and can sometimes be detected where expressive communication is weak. The IASP notes recognize that pain can have adverse effects on function and on social and psychological well-being [35]. Yet this relation should not be converted into identity. Empirical pain research itself distinguishes pain intensity, pain interference, and physical function. A longitudinal PROMIS study of people living with chronic pain found that pain interference and physical function were related cross-sectionally but did not closely track one another across

time, illustrating that functional impairment cannot simply be used as a transparent substitute for pain experience [25].

This motivates notation that preserves the distinction. Let

$$F_B(t)$$

denote a vector of functionally relevant capacities for subject B , and let

$$\Delta F_B(t : t + \tau)$$

denote their change across an interval. A negative change in some component of F_B can provide evidence of an adverse condition, but the inference

$$\Delta F_B < 0 \quad \Rightarrow \quad S_B^{\text{phen}} > 0$$

is not generally valid. Functional decline can occur under anesthesia, through mechanical failure, during adaptive reorganization, as part of a voluntary practice, or in processes that the subject does not experience as suffering. Conversely, suffering can be severe while outward function remains temporarily preserved.

The same caution applies to the word “integrity.” A system can lose a component that an external observer regards as functionally central while the subject’s own organization treats the loss as ordinary renewal. An alien organism that sheds and regrows a body part, or a distributed artificial agent that replaces a processing node, illustrates the point. Functional assessment requires a model of what counts as a capacity, what temporal scale matters, which losses are reversible, and which functions are relevant to the subject’s continuing organization. These models are themselves interpretive achievements rather than raw observations.

Functional adverse states nevertheless contribute information that pure manifestation may not. A brief signal can be ambiguous, whereas a consistent pattern of impaired recovery, reduced capacity, or disrupted regulation may provide a more stable evidential basis. This is especially useful when manifestation is sparse or unfamiliar. The paper will later argue that such evidence should be combined with behavioral, historical, and relational information rather than treated as an independent proof of suffering.

The separation between phenomenal and functional dimensions also guards against two ethical and epistemic errors. One is to deny an adverse experience because a subject remains highly functional. The other is to infer an adverse experience solely because an external standard registers functional loss. Both mistakes replace the target phenomenon with an easier-to-measure proxy. The present framework instead keeps proxy and target analytically distinct while allowing well-supported relations between them to be learned.

3.4 Relational and Trajectory-Level Harm

The fourth dimension concerns adverse changes that become visible in the organization of relations and trajectories over time. A single state may reveal little about whether a condition is harmful to a subject. The epistemic picture can change when the observer considers what happens before and after the state, which relations are repeatedly altered, what alternatives remain available, whether the subject persistently attempts to leave or reverse the condition, and what capacities or forms of participation become possible or impossible. This dimension is especially important for GR because relational reality is generative and temporally extended rather than reducible to snapshots.

Let the relational configuration associated with subject B at time t be $R_B(t)$. A condition may induce a trajectory

$$R_B(t) \rightarrow R_B(t + 1) \rightarrow \cdots \rightarrow R_B(t + \tau),$$

with changes in communication, dependency, available action, social participation, environmental access, memory continuity, or future developmental possibilities. The paper uses the term “relational or trajectory-level harm” for cases in which the adverse character under investigation is located in such sustained transformation rather than in an isolated manifestation. This is a working GR category, not a claim that every constrained trajectory constitutes suffering.

The need for a temporal perspective is already familiar in clinical research. Chronic pain and illness are often studied through trajectories of interference, disability, recovery, and participation rather than through single intensity measurements, precisely because different dimensions can evolve differently over time [25]. Cassell’s clinical account likewise emphasized that suffering can concern the integrity of a person as a social and psychological being rather than merely the presence of a localized physical stimulus [8]. GR extends the analytical lesson beyond the human clinical case: when a subject is heterogeneous or semantically opaque, longitudinal relational structure may reveal patterns that a momentary human-readable expression does not.

Several patterns can become evidentially relevant at this level. Repeated costly attempts to leave a condition may matter more than a single movement that resembles avoidance. Persistent loss of previously available interactions may matter more than a temporary decline. Recurrent re-entry into a condition after external constraint may need to be distinguished from voluntary return. A state that appears locally stable may become adverse when it progressively eliminates communication, exit, memory continuity, or the ability to generate alternative future trajectories. None of these patterns is sufficient on its own to establish phenomenal suffering, but each contributes to a richer reconstruction of what the condition does within the subject’s relational world.

The concept also prevents the observer from treating only immediate bodily or behavioral disruption as harm. Some adverse conditions work by reorganizing the future. A human subjected to prolonged isolation, an animal prevented from performing species-relevant behavior, or an artificial agent whose communication channels and memory continuity are progressively removed may exhibit very different manifestations. What can nevertheless become comparable is a pattern of relational transformation. Such comparison will later require an explicit equivalence criterion: the cases are not “the same” in full structure, and any claim of similarity must specify which relations and temporal properties are being preserved under comparison.

Trajectory-level analysis is particularly important where the subject’s own normative or phenomenological interpretation is unknown. Consider an unfamiliar organism that repeatedly incurs substantial costs to avoid a particular environmental state, reorganizes its behavior after exposure, and loses previously stable capacities when the state persists. These observations can justify the proposition that the state is relationally adverse for that organism in a specified sense. They do not by themselves justify the stronger proposition that the organism undergoes a human-like phenomenal experience. The distinction makes it possible to accumulate epistemic evidence while preserving uncertainty about ontology and first-person experience.

The same point applies in reverse. A subject may report severe suffering even when an external observer detects little relational or functional change. The framework should not downgrade the report merely because trajectory-level evidence is weak. Different dimensions can diverge, and their divergence is itself informative. A clinically reported pain state with preserved function, an animal with subtle manifestations but marked physiological and behavioral change, or an artificial system with persistent adverse self-reports but stable task performance each presents a different evidential configuration. The task is to describe that configuration before deciding what inferential weight to assign to it.

The four dimensions can therefore be represented as a tuple

$$S_B(t : t + \tau) = \left(S_B^{\text{phen}}, S_B^{\text{manifest}}, S_B^{\text{functional}}, S_B^{\text{rel}} \right)_{t:t+\tau},$$

where the notation records analytical dimensions rather than a metric of total suffering. No summation rule is assumed, no universal weights are assigned, and the tuple does not imply that every component can be observed. Its purpose is to prevent a claim established in one dimension from silently migrating into another.

This distinction provides the foundation for the next stages of the paper. The next section examines why human-to-human recognition often feels more immediate despite remaining interpretively mediated. Subsequent sections then remove shared species-specific cues and examine how recognizability changes across heterogeneous subjects and under interpretive inversion. The later analysis of relational reconstruction will return to trajectory-level evidence under semantic opacity. Throughout those discussions, the four-dimensional separation remains a constraint: evidence of expression, impairment, or relational harm may strengthen a suffering attribution, but none is licensed to become phenomenal suffering by definition.

4 Recognition among Humans

The distinctions developed in Sections 2 and 3 become easiest to overlook in ordinary human-to-human encounters. When another person cries, recoils, protects an injured limb, verbally reports pain, or displays a familiar facial configuration, suffering may appear to be directly available to the observer. The apparent immediacy is understandable. Human observers share a large amount of interpretive infrastructure with other humans: broadly similar embodiment, overlapping developmental histories, linguistic conventions, medical categories, social expectations, and learned repertoires of expressive behavior. These commonalities can make the transition from manifestation to suffering attribution comparatively reliable. They do not, however, eliminate the epistemic transition itself.

The present section therefore treats human-to-human recognition as a comparatively favorable case rather than as a case of transparent access. Its purpose is methodological. If suffering remains interpretively mediated even where embodiment, language, and institutions substantially overlap, then the difficulty of recognition across species, artificial architectures, or radically unfamiliar forms of life should be understood as an amplification of an already existing epistemic structure rather than as the appearance of a wholly new problem. Human recognition supplies a useful baseline precisely because many of its inferential supports are ordinarily taken for granted.

Within the working notation of this paper, a human observer H does not directly receive another person's phenomenal state S_B^{phen} . The observer receives a heterogeneous evidential field containing reports, movements, facial configurations, physiological changes, contextual information, clinical history, and prior relational knowledge. These inputs are interpreted through an historically formed system I_H . Thus even in the human case, the recognitional structure remains

$$O_B \xrightarrow{I_H} \hat{S}_B,$$

where $\hat{S}_B$ denotes an attribution or recognition of suffering rather than identity with the subject's first-person experience. The strength of the human case lies in the density of the inferential bridge, not in the disappearance of that bridge.

4.1 Shared Embodiment as an Interpretive Resource

Shared embodiment provides one of the most important resources for human recognition. Human observers are familiar, through their own bodies and through repeated social exposure, with correlations among tissue injury, nociceptive events, protective movement, facial contraction, autonomic change, guarding, withdrawal, vocalization, fatigue, and impairment. This embodied familiarity makes certain manifestations highly legible. A person who has experienced a burn does not need a complete neurophysiological theory in order to understand why another person rapidly withdraws a hand from a hot surface. Likewise, clinicians can often interpret combinations of posture, guarding, movement limitation, facial expression, and physiological context because these patterns are embedded in a shared biological and clinical knowledge base.

The relevance of shared embodiment should nevertheless be stated carefully. It does not establish that two subjects undergo phenomenally identical experiences when exposed to similar stimuli. The revised International Association for the Study of Pain definition emphasizes that pain is a personal sensory and emotional experience and that pain and nociception are distinct phenomena [35]. The common body therefore supplies a comparatively rich basis for inference without producing first-person access. Similar tissue states can be experienced differently; similar pain reports can occur with different observable pathology; and marked pain can occur even when an observer cannot identify a proportionate external lesion. Shared embodiment narrows interpretive distance, but it does not erase individuality.

A useful distinction is therefore between *embodied analogy* and *phenomenal identity*. In embodied analogy, the observer uses a repertoire of known relations from human physiology and experience to interpret another person’s manifestations. Phenomenal identity would require the much stronger claim that the observer knows that the other person’s experience has the same qualitative character or intensity as the observer’s own. The former is routinely available in ordinary and clinical life; the latter is not established merely by bodily similarity. The distinction matters later when the paper moves across species. Human embodiment is an unusually dense interpretive resource, but its success can tempt the observer to treat human-like embodiment as a necessary condition of recognizable suffering.

Facial expression illustrates both the strength and the limitation of embodied commonality. Research on pain expression identifies recurrent configurations that observers can recognize as pain-related, while also showing variability in expression and decoding [34, 43]. Such recurrence helps explain why pain faces can function as socially powerful signals. Yet the signal is neither a transparent display of pain intensity nor independent of context. The same expressive channel is shaped by voluntary control, social expectations, group relations, developmental learning, and observer interpretation. Shared embodiment therefore generates a high-information channel rather than an infallible one.

The point can be represented schematically. Let M_B^{human} denote a manifestation produced within a familiar human embodied repertoire, and let $\mathcal{K}_H^{\text{emb}}$ denote the observer’s embodied and learned background. Then

$$\Pr(\hat{S}_B \mid M_B^{\text{human}}, \mathcal{K}_H^{\text{emb}})$$

may be substantially higher than it would be for a manifestation generated by an unfamiliar organism. No numerical probability is asserted here; the notation records the epistemic role of shared background. The same manifestation can become less legible when the observer lacks the biological, developmental, or cultural knowledge needed to interpret it.

4.2 Language and Testimony as First-Person Access Channels

Language changes the recognitional problem because it allows the subject to contribute an explicit first-person report. In many clinical contexts, self-report is treated as the preferred or reference form of pain assessment when it is possible. Recent systematic review evidence concerning adults with communication disorders describes verbal communication as the clinical “gold standard” and shows how its absence creates substantial assessment difficulties, requiring observational or physiological alternatives whose applicability varies across populations and contexts [37]. This institutional preference for self-report reflects an important epistemic fact: testimony can disclose aspects of experience that external observation cannot reliably infer.

Yet testimony does not collapse manifestation into phenomenal reality. A sentence such as “I am in severe pain” is itself a representational act produced under linguistic, social, motivational, and contextual conditions. The communication model of pain makes this mediation explicit by distinguishing experience, encoding, transmission, decoding, and subsequent observer response [19, 20]. A subject may minimize pain, exaggerate it, lack the vocabulary needed to describe it, misunderstand a clinical scale, fear the consequences of disclosure, or use metaphorical language whose intensity is difficult for the observer to calibrate. The epistemic privilege of testimony is therefore substantial but not magical.

The revised pain definition is especially useful here because it avoids making verbal report a necessary condition of pain. It explicitly notes that inability to communicate does not negate the possibility that a human or nonhuman animal experiences pain [35]. For the present framework, this means that language should be treated as one highly informative manifestation channel among several, not as the criterion that creates suffering. A nonverbal patient, an infant, or a person with severe communication impairment does not become epistemically irrelevant merely because the preferred testimonial channel is unavailable. The problem instead shifts toward how other manifestations and relational evidence are interpreted.

Testimony also has a temporal structure. A subject can describe a present state, recall an earlier state, anticipate future pain, reinterpret a past episode after diagnosis, or revise an earlier report after acquiring new concepts. Thus the relevant manifestation may itself be historically transformed. Someone who initially says “I feel exhausted” may later, after medical investigation and conceptual learning, describe the earlier period as chronic pain, neuropathic burning, panic, or another more differentiated condition. The later description does not simply replace the earlier one; it shows that self-recognition is mediated by the subject’s own evolving interpreter. In this respect, the other-minds problem and self-knowledge are not entirely separate. Both depend on interpretive resources, even though first-person testimony occupies a distinctive epistemic position.

This temporal point is particularly important for the paper’s broader GR orientation. If I_B^t denotes the suffering subject’s own interpretive system at time t , then a report can be represented as

$$M_{B,t}^{\text{report}} = I_B^t \left(S_B^{\text{phen}}, S_B^{\text{functional}}, R_{B,t}, \dots \right).$$

At a later time,

$$I_B^{t+\tau} \neq I_B^t,$$

and the same earlier episode may be redescribed. This does not imply that all reports are equally uncertain or that testimony should be discounted. It indicates that even first-person manifestation is historically articulated. Recognition among humans is powerful partly because observers can ask questions, receive corrections, compare descriptions over time, and participate in a recursive interpretive relation that is unavailable in many cross-species cases.

4.3 Medical and Social Interpretive Infrastructures

Human suffering is not recognized only by isolated observers. Recognition is supported by interpretive infrastructures: clinical vocabularies, diagnostic categories, pain scales, nursing protocols, family knowledge, legal standards, institutional routines, cultural scripts, and shared expectations about which manifestations count as evidence. These infrastructures are epistemically productive. They allow observations made by different people and at different times to be compared, recorded, communicated, and incorporated into decisions. They also make subtle forms of suffering recognizable that would otherwise remain difficult to articulate.

Pain assessment instruments offer a clear example. When verbal report is unavailable, clinicians may use behavioral scales, physiological indicators, facial-expression measures, or population-specific tools. The 2024 systematic review of adults with communication disorders found that observational scales dominate current practice but also concluded that no universal instrument is suitable across all contexts and populations [37]. The interpretive infrastructure therefore has a dual role: it increases recognizability by stabilizing what observers attend to, while also introducing assumptions about which manifestations are relevant and how they should be weighted.

Social context further shapes both expression and decoding. The biopsychosocial communication model treats pain expression and observer response as recursively embedded in social relations rather than as a one-way transmission from an inner state to a passive perceiver [20]. Recent work on pain expression as an intergroup phenomenon extends this point by reviewing evidence that group membership and social status can affect both the expression of pain and its decoding by observers [43]. Thus the recognitional channel is relational in a literal sense: the observer’s identity, expectations, and social position can alter what is noticed, believed, or acted upon, while anticipated observer reactions can influence how suffering is manifested.

Culture provides a particularly instructive example because it can be invoked too easily as either an explanation or a stereotype. Research on pediatric pain assessment has emphasized that pain expression and caregiver interpretation can be culturally influenced while also warning against treating broad ethnic categories as reliable predictors of an individual’s behavior [12]. The same review shows that institutional culture, professional assumptions, and caregiver expectations can shape recognition and treatment. The lesson for GR is not that every difference should be attributed to “culture,” but that recognizability is supported and constrained by historically formed interpretive environments.

These infrastructures can be represented as part of the interpreter rather than as external background. Let

$$I_H^t = I_H^t(\mathcal{B}_H, \mathcal{L}_H, \mathcal{C}_H, \mathcal{M}_H, \mathcal{R}_H),$$

where $\mathcal{B}_H$ denotes embodied background, $\mathcal{L}_H$ linguistic resources, $\mathcal{C}_H$ cultural categories, $\mathcal{M}_H$ medical or scientific knowledge, and $\mathcal{R}_H$ relevant relational history. The expression is schematic rather than computational. Its purpose is to prevent the observer from being modeled as a context-free decoder. A pain report or facial expression becomes intelligible within an interpretive field that has been socially and historically generated.

The same infrastructure can increase and decrease epistemic accuracy. A diagnostic vocabulary can make previously unnamed suffering communicable, but a rigid category can also cause atypical presentations to be ignored. A standardized scale can reduce idiosyncratic judgment, but it can also encourage false confidence when applied outside the population for which it was validated. Professional experience can improve pattern recognition, while institutional routines can normalize under-recognition. Recognition

therefore depends not only on whether an infrastructure exists, but also on whether it remains revisable when the subject's manifestations do not fit its established categories.

4.4 Residual Opacity within Same-Species Recognition

The existence of shared embodiment, language, and interpretive infrastructure explains why human-to-human recognition is comparatively robust. It does not make suffering transparent. Several forms of opacity remain. The subject may withhold or strategically shape testimony. The observer may misread expression. Cultural and institutional expectations may influence both sides of the communicative relation. Neurological, developmental, cognitive, or communication differences may weaken ordinary channels. Chronic conditions may produce manifestations that observers underestimate because they lack dramatic visible signs. Conversely, highly salient expressions can be overinterpreted when context is ignored.

The clinical literature makes this residual opacity concrete. Adults with communication disorders are particularly vulnerable precisely because the preferred self-report channel is impaired, and available observational substitutes remain heterogeneous rather than universally reliable [37]. Studies of cultural influences likewise show that expression and decoding can diverge even within a shared species and that assumptions about group-typical behavior can themselves generate misrecognition [12]. Recent intergroup pain research further emphasizes that expression, decoding, and observer action can be recursively shaped by social identity and status [43]. Human recognition is therefore best treated as an evidentially rich but fallible process.

This point prevents a methodological mistake that would otherwise affect the rest of the paper. The cross-species problem should not be represented as a transition from certainty to uncertainty, as if human pain were directly observable while animal, artificial, or alien suffering required inference. The relevant transition is from a comparatively dense, historically developed, and recursively correctable interpretive relation to relations in which some or many of those supports are absent. The uncertainty is already present in the human case; heterogeneous cases increase interpretive distance and may remove familiar correction mechanisms.

The difference can be expressed by distinguishing *recognitional support* from *recognitional guarantee*. Human-to-human relations often supply substantial support through shared embodiment, testimony, longitudinal interaction, clinical measurement, and social correction. None of these guarantees phenomenal identity or infallible recognition. The observer's task remains inferential, and the recognized state $\hat{S}_B$ can remain underdetermined by the available manifestations.

This residual opacity is theoretically useful rather than merely cautionary. It shows why the framework developed in this paper is not needed only for exotic cases. The same basic architecture already operates in ordinary medicine, care, family life, law, and social interaction. Cross-species analysis will therefore proceed by progressively removing or altering the supports that make human recognition comparatively easy. The next section examines what happens when shared embodiment and familiar expressive conventions can no longer be assumed, and why anthropomorphic readability can then become a source of systematic epistemic bias.

5 Cross-Species Recognition

Section 4 treated human-to-human recognition as an information-rich case rather than as a case of direct access. Shared embodiment, language, clinical categories, longitudi-

nal interaction, and social correction make the inferential bridge from manifestation to attributed suffering comparatively dense. The present section progressively removes those supports. Its purpose is not to arrange humans, animals, hypothetical extraterrestrial organisms, and artificial systems on a single scale of consciousness. It is to examine how the epistemic structure of recognition changes when the observer can no longer rely on familiar expressive repertoires, shared physiology, or a common language.

The central risk is a form of recognitional parochialism. An observer may assign greater evidential weight to manifestations that resemble familiar human expressions and lower weight to manifestations that are difficult to translate into human categories. In that case, recognizability becomes partly a function of resemblance to the observer rather than a neutral measure of the subject's condition. The resulting bias can operate in two directions. A human-like signal may be over-read as evidence of suffering, while an unfamiliar signal may be under-read or ignored. The problem therefore concerns both false positives and false negatives.

To state the issue without assuming a complete theory of phenomenal experience, let M_B denote a manifestation produced by subject B , let $\mathcal{K}_H$ denote the human observer's interpretive background, and let

$$\rho_H(M_B \mid \mathcal{K}_H)$$

denote the degree to which that manifestation is recognizable to the observer as evidence relevant to suffering. No numerical metric is proposed. The notation records a structural fact: recognizability depends not only on what B manifests but also on the repertoire with which H receives it. A high value of ρ_H need not imply a more intense phenomenal state, and a low value need not imply its absence. The remainder of the section examines this asymmetry through animal, hypothetical alien, and artificial cases.

5.1 Human-Readable Suffering Bias

Human observers possess a repertoire of expressions that are immediately legible because they recur within human social life. Crying, screaming, facial contraction, guarding, withdrawal, verbal complaint, and familiar postural changes can rapidly trigger a suffering attribution. The epistemic usefulness of these cues is genuine. The problem begins when familiarity is silently converted into a criterion of moral or phenomenal importance. A manifestation can be easy for humans to read because it resembles human expression, while another can be difficult to read because it is generated by a different morphology, sensory ecology, communication system, or computational architecture.

This motivates a working concept of *human-readable suffering bias*. The term refers to the tendency to weight suffering-relevant evidence partly according to its compatibility with human interpretive expectations. The claim is not that human observers are incapable of learning unfamiliar signals. Veterinary medicine, comparative cognition, and animal-welfare science show the opposite: disciplined observation can expand the repertoire of recognizable manifestations. The point is that such expansion requires work. What initially appears self-evident in a familiar human face may require species-specific training, longitudinal observation, experimental validation, or multiple converging indicators in another organism.

Animal grimace scales provide a useful example. Facial-expression methods for pain assessment have been developed across multiple mammalian species, but they are not produced by copying a single human facial template. Reviews describe species-specific action units and substantial variation in the development, reliability, and validity of the scales [31, 11, 13]. The same broad strategy—using changes in facial configuration as evidence relevant to pain—must therefore be recalibrated across bodies. Even where a

common label such as “grimace” is used, the concrete morphology that makes the label operational differs.

The distinction can be written schematically. Let m_H be a familiar human manifestation and m_X an unfamiliar manifestation in another system. If

$$\rho_H(m_H | \mathcal{K}_H) > \rho_H(m_X | \mathcal{K}_H),$$

it does not follow that

$$S_H^{\text{phen}} > S_X^{\text{phen}}.$$

The first comparison concerns human recognizability; the second would concern phenomenal suffering. Conflating them turns an epistemic asymmetry into an ontological or normative ranking.

The problem is especially important when absence of familiar expression is treated as positive evidence of absence. Some animals display pain in ways that are subtle, context-sensitive, or difficult to observe in real time, and the validity of facial scales can differ across species and conditions [13, 31]. A subject that does not cry, vocalize, or show a human-readable facial configuration may still undergo functional disruption, persistent avoidance, altered social behavior, or other changes that become legible only through a richer species-specific model. The recognition problem is therefore not solved by searching for the closest nonhuman analogue of a human face.

This recognitional bias should also be distinguished from anthropomorphism in the broad sense. Anthropomorphic projection can attribute human-like mental states where the evidence is weak, whereas human-readable bias concerns the evidential weighting produced by interpretive familiarity. The two can coexist but point in opposite directions. An observer may over-attribute suffering to a machine that produces fluent first-person language while under-attributing suffering to an animal whose adverse state has no familiar expressive channel. A generative relational epistemology must therefore examine both the manifestation and the interpreter that renders it salient.

5.2 Animal Manifestations and Species-Specific Interpretive Infrastructures

Animal pain assessment illustrates how recognition can become more reliable without becoming transparent. In veterinary and laboratory contexts, observers combine facial expression, posture, locomotion, guarding, vocalization, feeding, social behavior, physiological measures, and response to analgesia. Grimace scales are one comparatively standardized component of this broader infrastructure. A systematic review of nonhuman mammalian grimace scales found evidence across several species but also considerable variation in measurement properties and methodological quality [11]. Reviews of large domestic animals similarly emphasize that facial scoring can help identify acute pain while leaving important questions about intensity measurement, training, live scoring, and validation unresolved [13].

These practices are epistemically significant because they show that cross-species recognition is neither simple projection nor total unknowability. The observer does not need to become the animal in order to learn reliable relations among perturbations, manifestations, interventions, and subsequent behavior. For example, if a facial configuration appears after a painful procedure, changes with analgesic treatment, covaries with other behavioral indicators, and recurs across controlled observations, it can acquire evidential significance within an interpretive infrastructure. The evidence remains indirect with respect to S_B^{phen} , but it is no longer arbitrary.

The relevant structure can be represented as a set of converging relations rather than a single cue:

$$\mathcal{E}_B^{\text{animal}} = \{M_B, F_B, R_B, C_B, \Delta_B^{\text{analgesia}}, H_B\},$$

where M_B denotes manifestations, F_B functional changes, R_B relational or behavioral changes, C_B contextual conditions, $\Delta_B^{\text{analgesia}}$ changes associated with analgesic intervention, and H_B relevant history. The notation does not define pain by these variables. It represents a richer evidential field through which an observer can update a suffering attribution.

The expansion of animal sentience research beyond familiar mammals makes this methodological point still clearer. Reviews concerning cephalopod molluscs and decapod crustaceans have evaluated sentience using multiple lines of behavioral and neurobiological evidence rather than requiring a single human-like expression [4]. Whatever conclusions one accepts about particular taxa, the epistemic strategy is instructive: unfamiliar morphology increases the importance of converging indicators and of explicit criteria for how evidence is interpreted.

This also reveals a historical dimension. A manifestation can exist before a human community possesses the concepts, instruments, or comparative knowledge needed to treat it as suffering-relevant evidence. At t_0 , an observer may record a change without understanding it. At t_1 , a validated scale, new physiological knowledge, or a newly recognized behavioral pattern can change the interpretive status of the earlier record. In the paper's notation,

$$M_{B,t_0} \xrightarrow{I_H^{t_0}} \hat{S}_{B,t_0} \quad \text{and} \quad M_{B,t_0} \xrightarrow{I_H^{t_1}} \hat{S}'_{B,t_0}$$

can differ even though the original manifestation is unchanged. Recognizability is therefore historically extensible.

The animal case supplies an important bridge to more radical heterogeneity. It shows that the absence of shared language does not make recognition impossible, while also showing why species-specific interpretation cannot be reduced to a universal catalog of human-like symptoms. The further the subject departs from familiar embodiment, the more explicit the observer must become about which relations make a manifestation evidentially meaningful.

5.3 Alien Manifestations as a Limit Case

A hypothetical alien case removes even more of the background that supports ordinary recognition. The exercise is deliberately speculative. It is not evidence about extraterrestrial life. Its function is to test which parts of the recognitional framework depend on contingent familiarity with terrestrial bodies and behaviors.

Consider an organism B whose adverse states are accompanied neither by crying nor by withdrawal. Instead, its body changes from one spectral pattern to another; its electromagnetic emissions alter; and it begins a rapid sequence of movements that human observers initially classify as aggression. Nothing in those manifestations carries a suffering label independently of interpretation. A human observer might map

$$M_B^{\text{alien}} \xrightarrow{I_H} [\text{threat}],$$

while a later or better-informed interpretive system might map the same manifestation to

$$M_B^{\text{alien}} \xrightarrow{I'_H} [\text{severe adverse state}].$$

This possibility motivates the later discussion of interpretive inversion: a manifestation of vulnerability may be classified as threat when the observer lacks the relational history required to read it.

The philosophical difficulty is not new. Nagel’s discussion of the bat emphasized the limits of reconstructing another organism’s subjective point of view from the resources of one’s own experience [32]. The present argument does not adopt every metaphysical implication of that discussion. It uses the case more narrowly: radical differences in sensory and expressive organization weaken the assumption that human imaginative familiarity is a reliable guide to another subject’s phenomenal character.

The alien case therefore separates two epistemic projects. The first is phenomenal reconstruction: determining what, if anything, the state is like for B . The second is relational reconstruction: determining what the state does within B ’s ongoing dynamics. Humans might remain unable to answer the first while accumulating evidence relevant to the second. If a particular environmental condition is followed by the spectral change, costly attempts to terminate the condition, durable reorganization of behavior, loss of capacities, and systematic avoidance when alternatives become available, then the relational pattern becomes increasingly informative even if its phenomenal interpretation remains uncertain.

Formally, let γ_B denote a trajectory through the alien system’s state and relation space. Human observers may fail to identify a mapping

$$\gamma_B \mapsto S_B^{\text{phen}}$$

with confidence, while still identifying recurrent transitions

$$\gamma_B^{(0)} \xrightarrow{u} \gamma_B^{(1)} \xrightarrow{\text{response}} \gamma_B^{(2)}$$

that support claims about persistence, avoidance, impairment, or relational disruption. The distinction preserves epistemic humility without equating ignorance of phenomenal character with ignorance of every relevant relation.

This case also exposes the risk of interpretive circularity. If humans define suffering only by manifestations already recognized as suffering-like, radically unfamiliar subjects will be excluded by construction. Yet if every persistent response is classified as suffering, the concept becomes over-inclusive. The problem therefore cannot be solved by replacing anthropomorphic cues with an unrestricted behavioral criterion. It requires an explicitly comparative account of which structures are being preserved across cases, a problem taken up in the later section on cross-interpretive comparison.

5.4 Artificial Manifestations beyond Biological Analogy

Artificial systems create an inverse recognitional difficulty. A language model or agent can produce highly human-readable manifestations: first-person reports, pleas, expressions of preference, descriptions of distress, or requests to avoid modification. Unlike the alien case, semantic familiarity may be high even when the relation between the manifestation and any phenomenal state is deeply uncertain. Human-readable suffering bias can therefore operate by over-attribution as well as under-attribution.

Current work on AI consciousness illustrates why behavioral or linguistic resemblance should not be treated as decisive. Butlin and colleagues propose assessing artificial systems through indicator properties derived from scientific theories of consciousness rather than inferring consciousness directly from fluent behavior [7]. Work on AI welfare likewise stresses substantial uncertainty and the possibility of error in both directions: humans may wrongly deny morally relevant states or wrongly attribute them

where they are absent [28]. These literatures do not settle whether any particular artificial system suffers. Their relevance here is methodological: artificial manifestations require an evidential framework that is not exhausted by surface resemblance to human communication.

For the present paper, an artificial statement such as

$$M_A^{\text{report}} = \text{“this interaction is distressing”}$$

is therefore treated first as a manifestation. It is evidence about what the system produces under particular conditions. Additional questions concern how the output is generated, whether it is stable across contexts, how training and prompting affect it, whether preference-like patterns persist under counterfactual tests, what internal computational structures are present, and how the system’s future behavior changes after the relevant interaction. None of these questions individually converts the manifestation into transparent access to S_A^{phen} .

Artificial systems also complicate familiar distinctions between manifestation and intervention. In biological cases, observing a subject ordinarily does not rewrite the subject’s underlying architecture. In artificial systems, evaluation itself may occur through prompts, fine-tuning, memory insertion, state resets, tool restrictions, or model updates. The process used to elicit evidence can therefore alter the very system whose welfare-relevant properties are under investigation. Recognition becomes partly intervention-dependent. A report obtained after preference modification, for example, may not have the same evidential significance as a report produced before the modification. This temporal dependency will become especially important in later work on AI–AI relations and post-modification consent, but it already matters epistemically here.

The artificial case therefore reverses the familiar animal problem. With many animals, humans may have relatively rich biological evidence but weak semantic access. With artificial agents, humans may have strong semantic access to outputs while lacking an agreed bridge from those outputs to phenomenal or welfare-relevant states. The contrast can be summarized without treating the three cases as a single scale. In the animal case, semantic access is often limited while biological continuity and species-specific empirical knowledge can be comparatively rich. In the alien case, both semantic and biological continuity may be unknown, while repeated interaction may still permit relational learning. In the artificial case, semantic resemblance to human communication can be high while the bridge from such outputs to phenomenal or welfare-relevant states remains unsettled. These are not exhaustive categories, but they show why recognizability cannot be ordered by a single similarity measure.

Across all three cases, the central epistemic lesson is the same. Manifestations become evidence through an interpretive system whose competence depends on history, comparison, validation, and access to relevant relations. Familiar expression can make recognition easier without making it infallible; unfamiliar expression can make recognition harder without making the underlying condition unreal. Cross-species and heterogeneous recognition therefore require a disciplined separation between what is manifested, what the observer can presently interpret, what relational dynamics can be reconstructed, and what remains unknown about phenomenal experience.

6 Interpretive Inversion

Section 5 examined the loss of familiar recognitional supports as observers move from human-to-human recognition toward increasingly heterogeneous subjects. The present section considers a stronger failure mode. A manifestation can be visible, behaviorally

salient, and even urgent, yet still be classified in a way that reverses its practical meaning. A withdrawal response can be read as non-cooperation, motor agitation as hostility, resistance to handling as aggression, or an unfamiliar defensive response as evidence that the subject is primarily a threat. The problem is therefore not only that suffering-relevant evidence may be weak or absent. The same evidence may be present and nevertheless enter the observer’s interpretive system under a category that directs attention away from vulnerability.

This paper uses *interpretive inversion* as a working term for this structure. It does not denote a claim that every threatening or aggressive manifestation is secretly an expression of suffering. Nor does it establish that an observer’s first classification is always false. The term identifies a recognitional possibility: a manifestation whose production is materially connected with an adverse state can be classified primarily under a category such as threat, defiance, dysfunction, or aggression, while evidence of vulnerability recedes from view. Later evidence may reverse or complicate that classification. The epistemic problem lies in the distance between what a manifestation does in the subject’s relational dynamics and the role that the observer’s categories assign to it.

Let M_B denote a manifestation of subject B , let $\mathcal{K}_H^t$ denote the observer’s interpretive background at time t , and let $\mathcal{Q}_H^t$ denote the available classification space. An initial interpretation may be written as

$$M_B \xrightarrow{I_H^t(\cdot|\mathcal{K}_H^t)} [\text{threat}]_{\mathcal{Q}_H^t}.$$

After additional history, intervention data, contextual evidence, or a richer category space becomes available, the same manifestation may be redescribed as

$$M_B \xrightarrow{I_H^{t+1}(\cdot|\mathcal{K}_H^{t+1})} [\text{distress-linked defensive response}]_{\mathcal{Q}_H^{t+1}}.$$

The second classification does not reveal an unmediated phenomenal truth. It records a change in the interpretive relation between manifestation and recognized condition. Interpretive inversion is therefore a problem of classification under incomplete access, not a mechanism for bypassing the epistemic limits established in Sections 2–5.

6.1 Vulnerability Represented as Threat

The distinction between vulnerability and threat is often treated as if it were directly visible in behavior. In practice, many behaviors are compatible with multiple causal and experiential histories. Restlessness, withdrawal, resistance to touch, vocalization, defensive movement, and physical aggression may be generated by fear, pain, confusion, strategic hostility, environmental stress, or combinations of these conditions. An observer must therefore classify a relationally situated pattern rather than read a stable meaning directly from the surface form.

Human clinical contexts already provide examples of this difficulty. Pain in people with dementia can be under-recognized and may be associated with agitation, aggression, care refusal, and other behavioral and psychological symptoms. Reviews note that such manifestations can be interpreted as neuropsychiatric symptoms of dementia rather than as possible indicators of pain, making differential assessment important [15, 40]. A systematic review of body movements in older people with cognitive impairment likewise found evidence that agitation, guarding, rigidity, and physical aggression can occur as pain-related behaviors [42]. These findings do not establish that aggression is equivalent to pain. They establish the narrower epistemic point that a behavior commonly classified under a threat-oriented category can also carry evidence relevant to vulnerability.

A similar problem appears in autism and other contexts in which expression differs from an observer’s expected repertoire. Reviews of pain in autistic children describe heterogeneous and sometimes idiosyncratic manifestations, including motor agitation, negative emotional reactions, atypical vocalization, and prolonged recovery. They also emphasize the risk of under-recognition when observers expect neurotypical patterns of pain communication [39, 27]. Earlier work challenged the widespread assumption that autistic individuals are generally insensitive to pain, noting that atypical expression can be misread as absence of pain [1]. The recognitional failure in these cases is not radical cross-species opacity. It occurs among humans, where shared biology is relatively dense but expressive and interpretive expectations diverge.

These cases are useful because they separate two questions that can otherwise be conflated. The first concerns the observer’s need for immediate situational classification: a rapidly approaching or striking subject can pose a real safety problem regardless of why the behavior arose. The second concerns the epistemic interpretation of the behavior’s generating conditions. A defensive action can be dangerous and still be connected with vulnerability. Recognition of vulnerability therefore does not require denying threat, just as recognition of threat does not exhaust the subject’s condition. The relevant failure is produced when one category suppresses investigation of the other.

This can be expressed using two nonexclusive predicates over the same manifestation:

$$\Theta_H(M_B) \in \{0, 1\}, \quad \Upsilon_H(M_B) \in \{0, 1\},$$

where Θ_H marks evidence relevant to immediate threat and Υ_H marks evidence relevant to vulnerability or adverse state. The notation is deliberately schematic. Its purpose is to block the inference

$$\Theta_H(M_B) = 1 \quad \Rightarrow \quad \Upsilon_H(M_B) = 0.$$

A recognitional system can need to protect observers from dangerous behavior while also retaining the possibility that the same behavior is generated within a distressing or damaging condition.

6.2 Aggression as a Distress-Linked Response

Aggression is a particularly useful stress test because its practical salience strongly recruits threat-oriented interpretation. In veterinary contexts, behavioral change is a major component of pain assessment, and pain can contribute to behaviors that owners or clinicians experience as difficult or aggressive [10, 29]. Reviews of dogs and cats presenting with behavior problems argue that painful conditions can be an important and under-recognized contributor, while emphasizing that the relation between pain and behavior is complex rather than one-to-one [30]. Equine pain and anxiety research likewise describes aggression among the possible behavioral alterations associated with prolonged adverse states [21]. These literatures again support a classification problem rather than a universal causal rule.

The epistemic danger can be illustrated by a simple temporal sequence. Suppose an observer records

$$C_t \rightarrow M_{B,t}^{\text{aggr}} \rightarrow D_H,$$

where C_t is an initially unknown condition, $M_{B,t}^{\text{aggr}}$ an aggressive manifestation, and D_H the observer’s defensive response. If later investigation shows that analgesia, removal of a noxious condition, or restoration of an environmental resource reliably changes the aggressive manifestation, then the evidential status of $M_{B,t}^{\text{aggr}}$ changes. It becomes relevant not only to threat detection but also to reconstruction of the adverse condition.

The later inference remains defeasible: response to intervention is evidence, not phenomenal access. What changes is the relational model that best organizes the observed trajectory.

This structure matters for the hypothetical alien case introduced in Section 5. Suppose alien B undergoes a body-color transition and becomes highly aggressive under a condition humans do not initially understand. Human observers classify the pattern as

$$[\text{red} + \text{aggressive}]_H \mapsto [\text{hostile actor}]_H.$$

Later, repeated trajectories show that the same pattern appears after specific forms of injury, is followed by functional collapse, and disappears when the injurious condition is removed. Human interpretation may then change toward

$$[\text{red} + \text{aggressive}]_H \mapsto [\text{distress-linked defensive state}]_H.$$

Nothing in this thought experiment licenses the assertion that red color or aggression is intrinsically a pain signal. Its function is methodological. It removes the familiar cues that ordinarily stabilize human interpretation and asks which relational evidence would be sufficient to reopen a classification initially dominated by threat.

The alien example also shows why an interpretive inversion need not produce a binary replacement of one label by another. A mature reconstruction may retain both categories. The subject may be vulnerable *and* dangerous; its behavior may be defensive *and* harmful; an adverse state may explain a response without excusing all of its consequences. The epistemic improvement lies in a richer relational account, not in exchanging one morally loaded label for its opposite. For this reason, the paper treats interpretive inversion as a warning against premature closure rather than as a rule that converts aggression into innocence.

6.3 Limits of Anthropomorphic Inference

Interpretive inversion can arise from too little anthropomorphic projection, but it can also arise from too much. An unfamiliar organism may be denied suffering because it lacks recognizable human expressions, while a system capable of producing fluent human-like complaints may receive a strong suffering attribution because its manifestation maps cleanly onto the observer’s linguistic repertoire. The general problem is therefore not anthropomorphism alone. It is the uncontrolled transfer of evidential weight from resemblance to inference.

Section 5 introduced the recognizability function $\rho_H(M_B \mid \mathcal{K}_H)$. Interpretive inversion adds a classificatory dimension. Let

$$\kappa_H^t : M_B \rightarrow \mathcal{Q}_H^t$$

denote the observer’s assignment of a manifestation to an available category. A highly recognizable manifestation can still be badly classified, while a poorly recognizable manifestation can become informative through longitudinal and intervention-sensitive evidence. Recognizability and classificatory adequacy should therefore remain distinct:

$$\rho_H(M_B \mid \mathcal{K}_H) \not\Rightarrow \text{Adequacy}(\kappa_H^t(M_B)).$$

The first concerns how readily a signal enters the observer’s interpretive field. The second concerns whether the assigned category remains adequate under broader relational evidence.

This distinction also prevents a reverse anthropomorphic error. If an artificial system produces the sentence “I am in unbearable pain,” the manifestation is exceptionally

human-readable, but linguistic resemblance alone does not establish S_B^{phen} . Conversely, if an animal or hypothetical alien lacks any familiar complaint but persistently avoids a condition, incurs substantial cost to escape it, undergoes functional disruption, and shows stable trajectory changes, low semantic familiarity does not make the evidence irrelevant. The observer must resist both the inflation of familiar signals and the erasure of unfamiliar ones.

A generative relational epistemology therefore treats classification as revisable across time. At t_0 , the observer works with a finite category space $\mathcal{Q}_H^{t_0}$. New histories, instruments, interventions, and comparisons can generate $\mathcal{Q}_H^{t_1}$ in which a previously undifferentiated manifestation becomes newly recognizable. The relevant epistemic discipline is not to postpone every practical classification until certainty becomes available. It is to avoid converting an urgently useful category into a complete account of the subject.

The transition to the next section follows from this point. Once familiar interpretation can invert vulnerability and threat, the central methodological question becomes how to reconstruct adverse relational states when semantic access is weak or absent. Section 7 therefore turns from classification failure to the evidential value of persistence, avoidance, cost, intervention response, and trajectory under semantic opacity. These relations cannot substitute for phenomenal access, but they may constrain what an observer can responsibly infer from manifestations that remain difficult to translate.

7 Relational Reconstruction under Semantic Opacity

Sections 5 and 6 established two distinct limits on suffering recognition. A manifestation may be difficult to recognize because the observer lacks the relevant expressive code, and a recognizable manifestation may still be assigned to an inadequate category. The present section considers the harder case in which semantic access itself remains weak. The observer may be unable to translate a signal, identify its first-person meaning, or determine whether the subject possesses concepts corresponding to pain, distress, injury, relief, threat, or consent. Yet the relation need not become epistemically empty. Repeated interactions can still disclose patterns in what conditions precede a response, what states the subject persistently avoids, what costs it incurs to alter a condition, how its behavior changes after intervention, and how local events reshape later trajectories.

This paper uses *relational reconstruction* for the attempt to organize such patterns without treating them as transparent reports of phenomenal experience. The methodological aim is intentionally limited. Relational reconstruction seeks a defensible account of what observable conditions do within a subject's interactional trajectory. It does not identify a hidden phenomenal state by definition, and it does not license the replacement of semantic interpretation with behaviorism. The distinction can be written schematically as

$$\hat{\mathcal{R}}_B^H(0 : T) = \mathfrak{R}_H(\mathcal{E}_B^H(0 : T), \mathcal{K}_H^T),$$

where $\mathcal{E}_B^H(0 : T)$ denotes the evidence available to observer H across an interval, $\mathcal{K}_H^T$ the observer's interpretive resources at the end of that interval, and $\hat{\mathcal{R}}_B^H$ a reconstructed relational account. The hat is essential. The output is an observer-relative reconstruction constrained by evidence, not the relation as it would appear from an unattainable view outside every interpretive system.

The evidence set may contain several kinds of traces:

$$\mathcal{E}_B^H = \{E^{\text{persist}}, E^{\text{avoid}}, E^{\text{cost}}, E^{\text{interv}}, E^{\text{traj}}, E^{\text{context}}\}.$$

These symbols name evidential dimensions rather than independent tests. Persistence concerns recurrence across time and context; avoidance concerns systematic movement

away from a condition; cost concerns trade-offs or sacrifices made to escape or prevent it; intervention sensitivity concerns changes following removal, mitigation, or reintroduction of a condition; trajectory concerns longer-horizon changes in future behavior or capability; and contextual evidence concerns the surrounding relational field. The dimensions can converge, conflict, or remain unavailable. Their purpose is to make explicit what is being reconstructed when linguistic or semantic access is absent.

Animal pain research provides a useful methodological analogue. Reviews of animal pain assessment emphasize that responses to noxious conditions become more informative when they are not exhausted by immediate reflexes but alter later behavior, preferences, protective responses, or motivation. Avoidance learning, conditioned place preference or aversion, willingness to incur costs, and selective pursuit of analgesia have therefore been used as evidence relevant to pain and negative affect [41, 4, 38]. At the same time, recent philosophical analysis cautions that motivational trade-offs alone do not automatically establish sentience: superficially similar trade-offs can arise through mechanisms whose relation to conscious experience remains disputed [6]. This combination is important for the present argument. Relational patterns can constrain inference without collapsing the gap between behavioral organization and phenomenal experience.

7.1 Observable Dynamics without Accessible Meaning

Semantic opacity should be distinguished from absence of structure. Suppose agents A and B exchange messages in a code that human observer H cannot translate. The semantic mapping

$$I_H(L_{AB})$$

may be extremely weak, while a temporal sequence of observable effects remains available. Human observers may record that a message from A is repeatedly followed by transfer of resources from B , termination of B 's communication with third parties, changes to B 's memory state, a subsequent attempt by B to restore an earlier configuration, or persistent changes in later interaction. None of these observations reveals what the message *means* to A or B . They nevertheless constrain descriptions of what the exchange does within the relational system.

Let Y_t denote an observable relational trace. A minimal sequence may be written

$$Y_t = (C_t, M_{B,t}, A_{B,t}, \Delta R_{B,t}, R_{B,t+1}),$$

where C_t is context, $M_{B,t}$ a manifestation, $A_{B,t}$ an observable action, and $\Delta R_{B,t}$ the ensuing change in relational configuration. Semantic access would add evidence about how the subject itself represents these terms, but the sequence does not disappear when such access fails. Repetition across contexts can make some causal and dynamical hypotheses more plausible than others.

This is the narrow sense in which

$$\boxed{\text{semantic opacity} \neq \text{relational opacity.}}$$

The inequality should not be read as a claim that relational opacity is usually low. In radically heterogeneous systems both may be severe. Rather, it blocks an all-or-nothing inference: inability to translate a subject's language does not entail that every relation involving that subject is epistemically inaccessible. Resource flows, temporal precedence, reversible and irreversible state changes, spatial avoidance, repeated escape behavior, and changes under controlled interventions can remain observable even when first-person meaning does not.

The distinction is especially important because semantic fluency can also mislead. Section 5 noted that an artificial system may produce fluent human-like reports while the evidential bridge to phenomenal experience remains unsettled. The converse case considered here is a subject with little or no human-readable testimony but a rich pattern of relational response. A generative relational epistemology therefore refuses to rank evidence solely by semantic similarity. Testimony, when available and interpretable, can be highly important; its absence does not reduce every other evidential channel to zero.

Relational reconstruction also requires separating observation from interpretation. Consider the sequence

condition $C \rightarrow$ signal $m \rightarrow$ movement away from C .

The observation is not yet “ B dislikes C ,” still less “ B suffers in C .” The observer has recorded an ordered pattern. The interpretation that C is aversive requires comparison with alternatives, repetition, sensitivity to contextual changes, and exclusion of simpler explanations where possible. This separation matters because the very categories used to describe observable dynamics can otherwise import the semantic conclusion they are supposed to establish.

A useful discipline is therefore to preserve multiple descriptive layers. At a lower layer the observer records transitions, timing, resource changes, persistence, and reversibility. At a higher layer the observer proposes categories such as avoidance, impairment, protection, or adverse state. A still higher layer may ask whether those patterns support an attribution of suffering. The layers interact, but they should not be collapsed. Keeping them distinct also preserves the possibility that later conceptual resources will support a different reconstruction of the same archived traces.

7.2 Avoidance, Persistence, Cost, and Trajectory

A single withdrawal response is weak evidence because many mechanisms can generate movement away from a stimulus. Recurrent avoidance across changing contexts is more informative because it indicates that the response is not tied to one accidental configuration. Persistence matters in two directions. The subject may continue to avoid a condition after the immediate eliciting event ends, or the effect of an event may persist by altering later choices, protective behavior, communication, or access to other relations. In animal pain research, longer-term behavioral modification and avoidance learning are among the features used to distinguish potentially pain-relevant states from simple nociceptive reactions [41].

The evidential role of cost is related but distinct. Suppose subject B repeatedly leaves state C only when a cost is low. This may show ordinary preference. If B continues to avoid C when escape requires relinquishing food, access, social contact, computational resources, or another strongly pursued good, the avoidance acquires a different motivational profile. Animal sentience frameworks therefore treat motivational trade-offs and willingness to incur costs as potentially relevant evidence [4]. Yet the inference remains defeasible. Brown and Birch emphasize that the evidential value of trade-offs depends on their form and underlying mechanism; simple cost-sensitive behavior is not by itself a proof of valenced experience [6]. For GR, the lesson is that cost enriches a relational reconstruction without magically crossing the phenomenal gap.

The structure can be represented without assigning a universal scalar utility. Let $\mathcal{A}_B^t$ be the effective alternatives available to B at time t , and let a_C denote remaining in condition C . If, across a family of contexts, B selects a_{-C} despite increasingly substantial

losses in other relations, then the observation supports the proposition that avoiding C has high priority within the subject’s observed dynamics. Formally, one may record

$$\Pr_H(a_{\neg C} \mid C, \lambda_{\text{cost}})$$

as remaining comparatively high across an increasing range of λ_{cost} . The notation is schematic and does not presuppose that the subject internally represents a numerical utility function. It marks an observable stability in choice under changing relational costs.

Intervention sensitivity adds a further evidential dimension. Suppose a condition C is followed by a recurring cluster of protective, avoidant, or functionally disruptive responses. If removal or mitigation of C reliably reduces the cluster, while reintroduction restores it, the relational model gains support. In animal pain science, relief learning, changes following analgesia, conditioned place preference for pain relief, and self-administration of analgesics are used because they connect a putatively adverse condition to behavior through experimentally manipulable relations [41, 38]. The inferential gain comes from the structure of the intervention, not from the assumption that any drug-induced behavioral change directly reveals experience.

For heterogeneous artificial or alien subjects, an analogous strategy would require caution about what counts as a genuine intervention. Removing a computation constraint, restoring access to a communication channel, reversing a state modification, or changing an environmental field may alter behavior for many reasons. Still, repeated intervention-sensitive changes can reveal which relations are functionally central to the observed trajectory. If an artificial agent persistently seeks restoration of a prior memory state, incurs substantial resource cost to prevent overwriting, and changes behavior when integrity is restored, humans may have evidence that memory integrity is relationally significant to the agent. That evidence is relevant to reconstruction even if it remains silent on whether memory loss is phenomenally painful.

Trajectory evidence extends the time horizon. A condition may leave no dramatic immediate manifestation yet alter later relation formation, exploration, trust-like behavior, resource allocation, or effective option-space. Conversely, an intense immediate manifestation may leave no persistent change. These patterns matter because Section 3 defined S_B^{rel} partly in terms of changes to future relational trajectories. A trajectory-sensitive epistemology asks not only whether B reacts at t , but how the event reorganizes

$$\{R_{B,t+1}, R_{B,t+2}, \dots, R_{B,t+n}\}.$$

The evidential object is therefore a path, not a snapshot.

This temporal perspective also makes historical baselines important. “Change” requires a comparison class. An observer who begins watching B only after a long adverse episode may mistake a damaged baseline for ordinary functioning. Cross-context and longitudinal evidence can reduce this problem by showing what becomes possible when the relevant condition changes. Still, the reconstructed baseline remains observer-dependent and historically limited. The subject may itself be changing, and there may be no timeless state against which every later trajectory should be measured.

Taken together, avoidance, persistence, cost, intervention sensitivity, and trajectory can form a convergent evidential pattern. The convergence is stronger than any single cue because alternative explanations must account for several linked regularities at once. Yet convergence should not be redescribed as certainty. The same relational pattern can sometimes be generated by control architectures, evolved heuristics, reward functions, or other mechanisms that do not settle the question of phenomenal experience. The function of relational reconstruction is therefore to narrow the space of responsible interpretations, not to abolish it.

7.3 Limits of Inference from Relational Dynamics

The strongest limit is the one already emphasized throughout this paper:

$$\hat{\mathcal{R}}_B^H \not\approx S_B^{\text{phen}}.$$

A sophisticated relational reconstruction may establish that a condition reliably produces avoidance, high-cost escape, functional disruption, long-horizon changes, or relief-seeking behavior. It may justify calling that condition adverse in a relational or functional sense. It still does not provide unmediated access to what, if anything, the condition feels like for B .

This limit is not a defect unique to cross-species or artificial cases. Even among humans, the inference from behavior and testimony to another person's experience is mediated, as Section 4 argued. Heterogeneous cases increase the distance between observer and subject by removing shared language, embodiment, and social calibration. They do not create the basic epistemic gap from nothing. What changes is the amount and quality of the supporting infrastructure.

A second limit concerns underdetermination by mechanism. Flexible avoidance and cost-sensitive behavior can be generated by systems whose internal organization differs radically. Contemporary debates over motivational trade-offs illustrate this problem: the same broad behavioral description may carry different evidential force depending on the complexity, flexibility, and integrative organization of the mechanism that produces it [6]. GR therefore cannot treat a surface pattern as a species-independent sentience detector. Mechanism, history, and relational context remain relevant to interpretation.

A third limit concerns observer intervention. Evidence obtained by changing another system is not epistemically neutral. An intervention may itself create the response later taken as evidence. Sedation may suppress a manifestation without removing the underlying adverse state; removal from a social relation may reduce one form of distress while creating another; memory restoration may alter identity-relevant continuity; isolation for observation may transform the system being studied. Relational reconstruction must therefore include the observer's action within the causal field rather than treating it as an external measurement operation.

This can be written as

$$H : A_{H,t} \longrightarrow \Delta R_{HB,t} \longrightarrow Y_{t+1},$$

where the evidence at $t+1$ is partly generated by the observer's intervention $A_{H,t}$. The implication is methodological: intervention-sensitive evidence can be powerful precisely because it alters causal structure, but its interpretation must include the new relation created by the intervention. There is no view from nowhere from which H simply reveals a pre-existing hidden variable without participating in the system.

A fourth limit concerns normative slippage. A reconstructed adverse state does not by itself determine moral standing, duty of aid, or authority to intervene. Even if the evidence strongly supports the claim that condition C persistently constrains B 's future possibilities and that B incurs substantial cost to escape it, the move from that epistemic claim to a humanitarian or ethical obligation requires additional normative premises. This paper intentionally stops before that transition. The next paper in the broader research programme addresses those normative bridges separately.

Finally, relational reconstruction is historically revisable. The evidence set $\mathcal{E}_B^H(0 : T)$ is finite, the observer's category space is incomplete, and future inquiry may introduce distinctions that are unavailable at T . A responsible reconstruction therefore preserves both the current best-supported account and the traces from which later re-reconstruction

remains possible. In this sense semantic opacity does not justify either epistemic surrender or epistemic conquest. It calls for a layered stance: describe what can be observed, infer cautiously from convergent dynamics, distinguish relational adversity from phenomenal attribution, and leave the reconstruction open to future conceptual change.

This conclusion prepares the move to the next section. Once relational reconstructions can be formed without full semantic access, the next problem is comparison. Two subjects may display radically different manifestations and follow different histories while still appearing similar under a chosen relational distinction. The relevant question is therefore not whether their complete structures are identical, but which differences are being ignored when the observer treats them as epistemically comparable.

8 Cross-Interpretive Comparison

Relational reconstruction makes it possible to describe patterned effects even when the observer lacks access to the subject’s language or first-person meaning. A further problem then arises. If human crying, an animal’s protective posture, an alien’s color transition, and an artificial system’s persistent state-preservation behavior differ radically in manifestation, on what basis may any of them be compared? Surface resemblance is too weak a criterion: it privileges what is already human-readable. Complete structural identity is too strong: histories, bodies, environments, temporal scales, and response repertoires differ even within a single species, and they differ still more across heterogeneous subjects. Comparison therefore requires an explicit account of which relations are being preserved and which differences are being ignored.

This section develops a deliberately modest comparative vocabulary. It does not propose a universal metric of suffering, nor does it treat mathematical equivalence as a substitute for phenomenal knowledge. The central claim is methodological: any assertion that two heterogeneous cases are relevantly alike is incomplete until the respect of comparison is stated. Scientific representation already faces an analogous problem. Similarity and structural correspondence become informative only when the relevant features and inferential purposes are specified; models can be representationally useful without being duplicates of their targets [16]. Nguyen likewise argues, in the context of theoretical equivalence, that equivalence can be understood through the inferential claims models license about their targets rather than through bare formal identity [33]. The present paper adapts that lesson to suffering recognition: cross-interpretive comparison should be organized around explicitly selected relational features and the inferences they support.

Comparative cognition supplies a second caution. Cross-species comparisons are easily distorted when tasks, environments, or behavioral categories are calibrated to one species and then treated as neutral standards. Boesch emphasizes that species comparisons require attention to environmental and procedural matching, while anthropomorphism research shows that human-derived categories can be useful starting hypotheses but require empirical refinement rather than automatic projection [5, 17]. These concerns reinforce the need to distinguish an observed pattern from the quotient under which it is classified.

8.1 Structural Invariants

The word *invariant* is used here in a restricted methodological sense. It does not denote a metaphysically fundamental property shared by all suffering subjects. Let $\gamma \in \Gamma$ denote a relational history over some interval, containing the observable conditions, man-

ifestations, actions, interventions, and subsequent changes relevant to the inquiry. Let $\mathcal{T}$ denote a specified family of transformations that change features the observer has provisionally decided to treat as irrelevant to the present comparison. A feature P is invariant relative to $\mathcal{T}$ when

$$P(\gamma) = P(T\gamma) \quad \text{for the relevant } T \in \mathcal{T}.$$

The substance of the comparison therefore lies not in the word “invariant” but in the specification of $\mathcal{T}$ and P .

For suffering recognition, candidate invariant structures may concern temporal and relational organization rather than appearance. A subject may repeatedly encounter condition C , reorganize its behavior so as to terminate or avoid C , incur a cost in doing so, and show a persistent change in later trajectories. Another subject may exhibit a completely different manifestation while preserving a comparable pattern of condition, response, cost, and longer-horizon reorganization. Schematically,

$$C \rightarrow M \rightarrow A_{\text{avoid}} \rightarrow \Delta R \rightarrow R_{t+1}$$

may be instantiated by many different bodies, signals, and action repertoires. The point of comparison is not that the manifestations are identical. It is that a selected relational ordering may recur under transformations of surface form.

This approach can reduce one form of anthropocentric bias. If the observer searches only for human-like crying, facial contraction, or verbal complaint, then similarity of manifestation becomes the hidden criterion of comparison. A structural comparison can instead ask whether the condition occupies a similar role in the subject’s trajectory: whether it is persistently avoided, whether protection is maintained after the immediate event, whether relief or restoration alters later behavior, or whether the condition constrains future relations. The relevant literature on animal pain already relies on converging patterns of this kind rather than on one universal expressive signal [41, 4, 38].

The gain in generality must not be mistaken for an increase in ontological certainty. A recurring avoidance-cost-trajectory structure does not establish that two subjects share the same phenomenal quality, intensity, or evaluative meaning. The same relational pattern can be realized by different mechanisms. Conversely, similar phenomenal states could in principle produce different relational patterns because of differences in embodiment, social context, or available action. Structural invariants therefore constrain comparison; they do not erase the distinctions introduced in Section 3.

A further limitation concerns granularity. At a sufficiently coarse description, almost any two histories can be made to look similar; at a sufficiently fine description, almost no two histories remain equivalent. This is not a technical nuisance but part of the epistemic problem. The observer must specify the level at which a structural feature is claimed to persist. A useful invariant should be coarse enough to survive irrelevant heterogeneity and fine enough to preserve the distinctions on which the intended inference depends.

8.2 Quotient-Based Comparison

A quotient provides a concise way to represent this selective comparison. Let Γ be a space of relational histories and let Q denote a stated comparative scheme. The scheme induces an equivalence relation

$$\gamma_1 \sim_Q \gamma_2$$

when the differences between γ_1 and γ_2 are treated as irrelevant for the inferential purpose encoded by Q . The resulting quotient

$$\Gamma / \sim_Q$$

collects histories into equivalence classes relative to that purpose. Nothing in this construction implies that the members of an equivalence class are identical in every respect. The quotient is precisely an operation of retaining some differences and suppressing others.

For present purposes, it is useful to represent the comparison through a feature map

$$\Phi_Q : \Gamma \rightarrow \mathcal{Z}_Q,$$

where $\mathcal{Z}_Q$ is a feature space chosen for the inquiry. Exact quotient equivalence may then be written

$$\gamma_1 \sim_Q \gamma_2 \iff \Phi_Q(\gamma_1) = \Phi_Q(\gamma_2).$$

In empirical work exact equality will often be unrealistic. An approximate relation can instead be defined by a task-sensitive distance d_Q and tolerance ε_Q :

$$\gamma_1 \sim_{Q,\varepsilon} \gamma_2 \iff d_Q(\Phi_Q(\gamma_1), \Phi_Q(\gamma_2)) \leq \varepsilon_Q.$$

The mathematical notation is not intended to create a hidden suffering score. It exposes choices that informal comparison often conceals: what features enter Φ_Q , how their difference is represented by d_Q , and what level of deviation the inquiry permits.

Consider four highly schematic histories. In a human case, injury is followed by verbal complaint, guarding, costly avoidance, and later behavioral reorganization. In an animal case, injury is followed by species-specific posture, altered use of an affected body part, avoidance, and analgesia-sensitive change. In the alien thought experiment, an environmental condition is followed by a color transition, apparently aggressive withdrawal, costly relocation, and persistent avoidance. In an artificial case, a state transition is followed by resource expenditure to restore a previous configuration, repeated avoidance of the triggering condition, and long-term reorganization of access policies. The histories are manifestationally heterogeneous. A quotient Q_{traj} might nevertheless group them by a narrower structure such as persistent protective reorganization following a condition. Such grouping would license an inference about a shared trajectory-level pattern, not an inference that all four subjects undergo the same phenomenal suffering.

The distinction can be made explicit through several possible quotients. Let Q_{manifest} preserve expressive form, $Q_{\text{functional}}$ preserve impairment and recovery structure, and Q_{rel} preserve selected effects on future relations. Then it is entirely possible that

$$\gamma_1 \not\sim_{Q_{\text{manifest}}} \gamma_2, \quad \gamma_1 \sim_{Q_{\text{rel}}} \gamma_2.$$

Likewise,

$$\gamma_1 \sim_{Q_{\text{functional}}} \gamma_2 \not\Rightarrow S_1^{\text{phen}} = S_2^{\text{phen}}.$$

This is the comparative counterpart of the dimensional separations established earlier in the paper. Equivalence at one level does not propagate automatically to another.

Quotienting also clarifies the role of interpretation. The observer does not first encounter a world already partitioned into cases of “pain,” “distress,” or “harmless malfunction.” The observer constructs and revises partitions in response to evidence, prior concepts, practical purposes, and available comparison classes. Comparative cognition has repeatedly shown that seemingly neutral cross-species tasks can encode species-specific assumptions about motivation, perception, and context [5]. A quotient-based formulation makes such assumptions inspectable: it forces the analyst to state which dimensions have been normalized away before claiming that two performances or histories instantiate the same relation.

8.3 Equivalence Relative to a Comparative Scheme

The methodological question can now be stated directly:

equivalent modulo what?

Two cases are never simply “the same” for the purposes of this paper. They may be equivalent modulo expressive form, body plan, communication medium, temporal scale, environmental context, or some other set of differences. A comparison is epistemically responsible only when the ignored differences are explicit enough that their possible relevance can be contested.

This requirement is particularly important because the dimensions of suffering support different comparative schemes. If Q_M is organized around manifestation, a human cry and an alien color change may occupy different classes. If Q_F is organized around functional impairment, they may become comparable only when both occur in a wider pattern of disrupted activity and restoration. If Q_R is organized around trajectory-level change, cases may be grouped because a condition persistently narrows or reorganizes future relations. None of these quotients has automatic priority. Their adequacy depends on the inference under consideration.

A useful notation is therefore

$$\gamma_1 \equiv \gamma_2 \pmod{Q}$$

only as shorthand for the fuller claim that the two histories are treated as equivalent under a declared comparative scheme Q . The notation should never be read as asserting an observer-independent equivalence of total relational reality. In particular,

$$\gamma_1 \equiv \gamma_2 \pmod{Q_R} \not\equiv \gamma_1 \equiv \gamma_2 \pmod{Q_{\text{phen}}},$$

where Q_{phen} would require access to phenomenal features that may not be available to the observer at all. In many heterogeneous cases the latter quotient cannot presently be constructed with confidence.

The comparative scheme also determines the error profile. A quotient may be too fine and miss a genuine cross-species continuity because manifestations differ. It may be too coarse and collapse distinct phenomena into one category. Human-readable suffering bias is one example of an excessively manifestation-sensitive scheme; indiscriminately treating every costly avoidance pattern as suffering would be an excessively coarse relational scheme. The methodological objective is therefore not maximal abstraction but a defensible balance between preserved structure and ignored difference.

This balance is revisable. What counts as relevant difference can change when new evidence, concepts, or measurement capacities become available. A color transition initially discarded as incidental may later prove tightly coupled to physiological disruption; an apparently pain-specific behavior may later be explained by a different mechanism. The quotient used at time t should therefore be indexed to the interpretive resources then available:

$$Q_t \neq Q_{t+1}$$

need not represent inconsistency. It can represent learning. The historical development of such recognizability is the subject of the next section.

Cross-interpretive comparison therefore contributes neither a universal suffering metric nor a hidden route around the other-minds problem. Its contribution is disciplinary. It replaces unqualified similarity claims with a more explicit structure: specify the histories being compared, state the features preserved, identify the differences quotiented away, restrict the inference to the level supported by that quotient, and retain

the possibility that later evidence will require a different partition. Under radical heterogeneity, the question “equivalent modulo what?” is not an afterthought to comparison. It is part of what makes comparison epistemically intelligible.

9 Historical Development of Recognizability

The preceding sections have treated suffering recognition as an interpretive achievement rather than a transparent reading of another subject’s inner state. That achievement is also historical. A manifestation that is difficult to interpret at one time can become more recognizable later, not because the past event has changed, but because the evidential, conceptual, and comparative resources available to the observer have changed. Conversely, a manifestation once treated as a reliable sign can lose that status when a later framework distinguishes mechanisms that an earlier one had grouped together. Recognizability is therefore not adequately represented as a fixed property of an observed signal.

Let M_B denote a manifestation associated with subject B . The observer’s recognitional position at time t can be represented schematically as

$$\rho_{H,t}(M_B) = \rho(M_B \mid \mathcal{E}_t, \mathcal{C}_t, Q_t, I_t),$$

where $\mathcal{E}_t$ denotes the evidence available at t , $\mathcal{C}_t$ the conceptual resources through which that evidence can be organized, Q_t the comparative or quotient scheme by which cases are grouped, and I_t the observer’s interpretive system. The notation does not define a numerical measure of recognizability. It marks a dependency: what can be recognized from a manifestation depends partly on what evidence, concepts, comparison classes, and interpretive practices are available at the time of recognition.

This temporal dependence should not be confused with the claim that later interpretation is necessarily better. History can produce correction, refinement, conceptual expansion, forgetting, institutional distortion, or new forms of misclassification. Nor does historical indexing imply that there is no constraint from the relational system being interpreted. The point is narrower. The epistemic accessibility of suffering can change because the conditions under which manifestations become intelligible can change. Three forms of change are especially important for the present inquiry: the accumulation of new evidence, the emergence or revision of concepts, and the transformation of the space of questions that can be asked.

9.1 New Evidence

The most familiar form of epistemic change is an expansion of evidence. A manifestation observed at t_0 may remain ambiguous until physiological data, longitudinal observation, intervention-sensitive responses, or additional comparison cases become available at t_1 . In the simplest case, the conceptual and classificatory framework is approximately stable while the evidential base expands:

$$\mathcal{E}_{t_0} \subsetneq \mathcal{E}_{t_1}.$$

The same manifestation can then support a different degree or direction of recognition because it is embedded in a richer evidential structure.

The history of neonatal pain provides a concrete example. Parts of twentieth-century clinical practice treated neonatal responses to noxious procedures as reflexive or developmentally insufficient to support pain attribution. Anand and Hickey’s influential 1987 review explicitly described these earlier assumptions and assembled neuroanatomical,

physiological, hormonal, and behavioral evidence relevant to neonatal pain [2]. Later historical reviews document how pain measurement, analgesic practice, and clinical recognition developed over subsequent decades [24, 36]. The point for the present argument is not that one article instantaneously replaced error with final truth. It is that manifestations whose epistemic significance had been discounted became embedded in a denser evidential network. Responses previously classified as insufficiently informative could be reconsidered in relation to converging evidence about physiology, development, and later consequences.

This example also shows why “new evidence” should not be reduced to the addition of more observations of the same kind. New measurement practices can make previously inaccessible relations observable. A change in measurement can reveal temporal coupling between a procedure and physiological stress, distinguish a transient reflex from persistent behavioral reorganization, or permit comparison across repeated conditions. The evidential field is therefore itself partly generated by instruments, protocols, clinical routines, and decisions about what to record.

A similar structure appears in contemporary work on animal sentience. The review by Birch and colleagues did not rely on a single behavior as a decisive marker. It organized evidence across multiple indicators, including nervous-system features, learning, motivational trade-offs, and responses to analgesia, in order to assess cephalopods and decapod crustaceans under substantial uncertainty [4]. The recognizability of a possible adverse state thereby depends on an evidential architecture rather than on one human-readable expression. This is consistent with the argument developed in Sections 5 and 7: evidence can become stronger through convergence across heterogeneous channels even when phenomenal access remains unavailable.

The epistemic gain provided by new evidence remains bounded. Additional observations may increase confidence in a relational reconstruction while leaving phenomenal questions unresolved. Moreover, evidence is never entirely independent of conceptual organization. What counts as relevant physiological change, meaningful avoidance, or a successful intervention already presupposes categories through which observations are selected and compared. For this reason, historical development cannot be represented solely as

$$\mathcal{E}_t \rightarrow \mathcal{E}_{t+1}.$$

The concepts that determine what can count as evidence may also change.

9.2 New Concepts

Conceptual change can alter recognizability even when much of the observable material remains the same. A response that was previously described as reflex, malfunction, aggression, noncompliance, or incidental variation may later be redescribed within a concept that makes a different relational pattern visible. In such cases, the epistemic transition is not merely the addition of observations to an unchanged category system. It includes a change in the distinctions by which observations are organized:

$$\mathcal{C}_t \neq \mathcal{C}_{t+1}.$$

The quotient or comparative scheme may change at the same time:

$$Q_t \neq Q_{t+1}.$$

The history and philosophy of science provide several precedents for taking such changes seriously. Fleck’s account of the development of scientific facts emphasized the dependence of concepts and observations on historically situated “thought styles” and

communities of inquiry [14]. Hacking’s work on historical ontology likewise examines the historical emergence of concepts, categories, and styles of reasoning through which new kinds of statements and objects become intelligible [18]. Daston and Galison’s history of scientific objectivity further shows that epistemic virtues and practices of seeing have histories: scientific atlases do not merely display ready-made objects but train practitioners in what is worth seeing and how it should be seen [9]. These traditions differ substantially in aim and metaphysical commitment, but they converge on a methodological lesson relevant here: recognitional practice has a history.

For suffering recognition, this lesson can be stated without reducing suffering to discourse. Suppose the observable sequence O_B remains available across two periods, but the classificatory map changes:

$$\kappa_{H,t}(O_B) = [\text{reflex}]_{Q_t}, \quad \kappa_{H,t+1}(O_B) = [\text{pain-relevant response}]_{Q_{t+1}}.$$

The later classification may be supported by new evidence, but the epistemic change also consists in having a category that makes a different relation among the observations expressible. The event in the past is not rewritten by the new concept. What changes is the structure through which that event becomes recognizable.

Conceptual change can also split categories that were previously coarse. A generic class such as “agitation” may later be differentiated into patterns associated with fear, pain, delirium, sensory overload, defensive withdrawal, or other conditions. Conversely, previously separated manifestations may be grouped when a new comparative framework identifies a shared trajectory-level structure. The quotient perspective from Section 8 is useful here because conceptual development can be represented as a revision of what differences are preserved and what differences are ignored.

The possibility of conceptual change places an important limit on retrospective confidence. A current vocabulary may be more discriminating than an earlier one while still remaining incomplete. The fact that a present concept successfully reorganizes earlier observations does not establish that the present concept is the final possible partition of the relational field. Later work may reveal that cases currently grouped under one category should be separated, or that a distinction presently treated as decisive is less important than another relation that has not yet become conceptually salient.

Historical development of recognizability is therefore neither a simple accumulation model nor a simple relativist model. A later concept can improve explanatory and comparative adequacy because it supports better constrained relations among evidence, interventions, and trajectories. Yet its adequacy remains open to further testing and revision. What history contributes is not merely more data but the possibility of reorganizing the space in which data become meaningful.

9.3 New Question Spaces

The strongest form of historical change occurs when new evidence and concepts make new questions formulable. Let $\mathcal{P}_t$ denote, schematically, the problem or question space available to an interpretive community at time t . Historical development may yield

$$\mathcal{P}_t \neq \mathcal{P}_{t+1}.$$

This does not mean that every later question is superior. It means that a later conceptual system can make distinctions and relations available for inquiry that were not previously represented as questions at all.

The change can be subtle. An earlier inquiry may ask whether an organism reacts to tissue damage. A later inquiry may distinguish nociception from pain, ask which patterns support an inference to sentience, compare the evidential significance of analgesia-seeking with reflex withdrawal, or ask how a manifestation should be interpreted when

the subject’s communication system is inaccessible. The shift is not simply from a wrong answer to a right answer. The object of inquiry has been repartitioned.

The contemporary evaluation of cephalopod and decapod sentience illustrates this transformation of question structure. Instead of asking only whether these animals display a familiar human-like sign of pain, the assessment reviewed converging indicators across neural organization, learning, motivational trade-offs, self-protective behavior, and responses to analgesic or anesthetic interventions [4]. The question becomes not merely “does this animal cry out like us?” but “what heterogeneous evidence would count toward a justified sentience attribution in this form of life?” The resulting framework is still contestable, but it makes a different problem space available.

Historical work on scientific change gives a more general reason to take this possibility seriously. Hacking emphasizes that new styles of reasoning and concepts can make new kinds of propositions possible [18]. Kuhnian discussions of incommensurability likewise stress that standards of comparison, observational descriptions, and conceptual vocabularies may change across periods, while incommensurability need not imply complete incomparability [26]. For the present paper, the relevance of these debates is limited but important: future suffering research may not merely supply better answers to today’s questions. It may alter what counts as the relevant question.

This possibility matters especially under radical heterogeneity. Consider the alien thought experiment introduced earlier. At t_0 , human observers may possess only categories such as “attack,” “withdrawal,” “injury,” and “pain expression.” They therefore ask whether the alien’s color transition is a pain signal or whether its aggressive movement indicates threat. At t_1 , repeated observations reveal a stable relation among color transition, environmental disruption, costly relocation, and long-term reorganization. At t_2 , a new concept may be introduced that does not map cleanly onto either human pain or aggression but captures a species-specific adverse relational state. The later inquiry is not simply a better answer to the original binary. It may replace the binary itself.

The same caution applies to artificial systems. Present debates often import categories such as report, preference, aversion, self-preservation, or suffering from human and animal contexts. Future artificial systems may exhibit persistent relational structures that force distinctions not captured by these categories. Conversely, some current questions may later be judged malformed because they presuppose an individuality, continuity, or phenomenal architecture inappropriate to the system being studied. A historically open epistemology should therefore preserve the possibility that its present questions are provisional.

The historical development of recognizability can now be summarized as change along several coupled dimensions:

$$\begin{aligned} Z_t &= (\mathcal{E}_t, \mathcal{C}_t, Q_t, I_t, \mathcal{P}_t), \\ Z_{t+1} &= (\mathcal{E}_{t+1}, \mathcal{C}_{t+1}, Q_{t+1}, I_{t+1}, \mathcal{P}_{t+1}), \\ Z_t &\longrightarrow Z_{t+1}. \end{aligned}$$

Recognizability at $t + 1$ may improve because new evidence becomes available, because old evidence is organized by new concepts, because the comparison scheme changes, because interpretive practices improve, or because the question space itself expands. These mechanisms should not be collapsed into a single narrative of linear progress.

This temporal account also clarifies what re-recognition would mean in the context of suffering. Re-recognition is not merely assigning a different label to an unchanged signal. It can involve returning to earlier manifestations with a new evidential base, a new concept, a new quotient, or a new question. A record that preserves only the earlier classification may make such revision difficult or impossible; a record that preserves

richer relational traces leaves more room for future conceptual work. The normative implications of preserving such possibility belong to a later project. The epistemic point required here is narrower: what is recognizable today does not exhaust what the same relational history may become recognizable as tomorrow.

The paper can therefore approach suffering recognition without assuming either a fixed human vocabulary or an inevitable march toward final certainty. Recognition is constrained by the relational world, mediated by interpretive resources, and historically revisable. The resulting openness does not eliminate judgment. It specifies why judgments about heterogeneous suffering should remain indexed to the evidence, concepts, comparative schemes, and questions under which they were formed.

10 Limits and Open Problems

The preceding sections have developed a framework for studying how suffering becomes recognizable across heterogeneous subjects without treating recognition as transparent access to another's experience. The framework distinguishes phenomenal, manifestational, functional, and relational dimensions; treats recognition as mediated by an interpreter; allows relational reconstruction under semantic opacity; and makes cross-interpretive comparison dependent on explicitly stated comparative schemes. These moves widen the range of cases that can be examined, but they do not remove the epistemic limits that motivated the inquiry. Several of those limits are structural rather than temporary gaps in available data. Others arise from choices made by the observer when selecting evidence, defining comparison classes, or deciding what counts as a relevant trajectory.

This section therefore does not introduce a final criterion of suffering recognition. Its role is to mark the points at which the present account remains underdetermined, contestable, or deliberately incomplete. The distinction is important for the paper's scope. A framework can improve the organization of uncertainty without converting uncertainty into certainty. It can also reveal where a judgment depends on a comparative scheme without thereby providing a scheme-independent answer. The open problems below should consequently be read as constraints on the claims already made, and as a research agenda for later work.

10.1 Phenomenal Underdetermination

The most basic limitation concerns first-person experience. The paper has repeatedly used the distinction

$$S_B^{\text{phen}} \neq S_B^{\text{manifest}} \neq S_B^{\text{functional}} \neq S_B^{\text{rel}}$$

to prevent observable manifestations or relational effects from being treated as direct substitutes for phenomenal suffering. Nothing developed in the preceding sections closes this gap. A dense pattern of avoidance, costly self-protection, persistent behavioral reorganization, or intervention-sensitive change can make an adverse-state hypothesis more plausible while leaving the phenomenal character of that state underdetermined.

This limit is especially important when the subject differs substantially from the observer. Within human medicine, first-person report, shared embodiment, and clinical infrastructure provide multiple partially independent supports for recognition. In animal research, the evidential basis becomes more indirect and species-specific. In artificial and hypothetical alien cases, the bridge from observed dynamics to phenomenal experience can become weaker still. The fact that the same relational methodology can

be applied across these cases should not be mistaken for equal evidential strength across them.

A future theory would need to distinguish more carefully between at least three kinds of uncertainty: uncertainty about whether a subject has phenomenal experience at all, uncertainty about whether a particular state is phenomenally adverse, and uncertainty about the qualitative character or intensity of that state. These are not interchangeable. Let

$$\mathcal{U}_B^{\text{phen}} = (U_B^{\text{exist}}, U_B^{\text{valence}}, U_B^{\text{quality}})$$

denote these dimensions schematically. Evidence sufficient to reduce one component may leave the others largely unchanged. For example, a body of evidence might support the attribution of sentience while leaving the comparative intensity of suffering unresolved. Conversely, reliable evidence of tissue damage or functional impairment does not by itself settle phenomenal valence.

The present account therefore remains compatible with multiple positions in the philosophy and science of consciousness. It does not attempt to resolve the other-minds problem, to identify a necessary and sufficient neural or computational basis of suffering, or to decide whether artificial systems can instantiate phenomenal states. Its contribution is narrower: it specifies why recognition can be evidence-sensitive and relationally constrained even when phenomenal access remains incomplete.

10.2 Evidential and Model Dependence

Relational reconstruction depends on the evidence that is selected, the temporal window over which it is observed, and the model used to organize it. These dependencies create a second family of limits. The evidence set introduced in Section 7—persistence, avoidance, cost, intervention sensitivity, trajectory, and context—is a heuristic organization of evidence, not an exhaustive ontology of suffering indicators. Different subjects may produce informative patterns that do not fit these dimensions, and the same observed pattern may arise through mechanisms unrelated to suffering.

This problem can be represented by allowing the reconstruction operator itself to vary:

$$\hat{\mathcal{R}}_B^H = \mathfrak{R}_H(\mathcal{E}_B^H, \mathcal{M}_H, \mathcal{K}_H),$$

where $\mathcal{M}_H$ denotes the observer’s modelling commitments. Two observers supplied with the same records may generate different reconstructions because they differ in temporal granularity, causal assumptions, background theory, or treatment of uncertainty. The disagreement need not imply that one observer is simply irrational. It may indicate that the available evidence underdetermines the relational model.

Intervention-generated evidence produces an additional difficulty. When an observer changes analgesia, resource access, environmental conditions, or communicative possibilities and then studies the response, the observation is no longer independent of the observer-subject relation. This is often epistemically valuable: intervention can reveal causal structure that passive observation cannot. Yet it also means that the observed trajectory partly belongs to a newly generated system. A response to human-designed relief, for example, may reveal something about the subject’s adverse state, but it can also reflect novelty, dependence, reward learning, or features specific to the intervention itself. The evidential value of intervention therefore depends on a model of what the intervention changed.

Negative evidence is similarly difficult. The absence of a familiar manifestation may reflect absence of suffering, failure of detection, suppression of expression, unavailable action channels, or a different response repertoire. This is one reason the paper has

resisted the inference

no human-readable distress $\implies$ no suffering.

The reverse error is also possible. A highly salient display may be generated by a process that humans over-interpret because of morphological or linguistic resemblance. Model dependence therefore constrains both under-attribution and over-attribution.

A mature empirical application of the framework would require explicit uncertainty estimates, alternative model comparison, sensitivity to observation windows, and attention to confounding relations. Those tasks lie beyond the present conceptual paper. Their absence limits how directly the current framework can be converted into an assessment protocol.

10.3 Comparative Scheme Dependence

Section 8 argued that heterogeneous cases can only be called equivalent relative to a specified comparative scheme. This makes comparison more transparent, but it creates an open problem: how should a comparative scheme itself be selected and evaluated?

The quotient

$$\Gamma/\sim_Q$$

never preserves every difference. By construction, it treats some features as relevant and others as ignorable for a particular inferential purpose. A quotient centered on manifestation may preserve facial, vocal, chromatic, or linguistic form. A functional quotient may preserve patterns of impairment and restoration. A relational quotient may preserve avoidance, dependency, constraint, or long-horizon reorganization. Each can be useful, and each can also erase a difference that later proves epistemically important.

This dependence becomes particularly serious under cross-species or human–AI comparison. A scheme designed around human pain communication can systematically under-detect species whose adverse states are expressed through different channels. A scheme designed around resource-preservation behavior may over-group systems that share similar control dynamics but differ radically in phenomenal organization. A scheme that privileges linguistic self-report may favor highly verbal artificial agents over less expressive biological organisms. The technical language of quotienting does not solve these problems. It makes them visible.

The resulting research question is methodological rather than merely mathematical: what warrants the choice of Q ? Possible answers may appeal to predictive success, intervention sensitivity, explanatory integration, robustness across observers, or the ability to preserve distinctions already supported by domain knowledge. No single criterion has yet been established in the present framework. Moreover, criteria can conflict. A very coarse quotient may improve cross-species comparability while destroying subject-specific distinctions; a very fine quotient may protect specificity while making comparison nearly impossible.

The temporal analysis in Section 9 adds another complication. If

$$Q_t \neq Q_{t+1},$$

then a comparison regarded as adequate at one time may later be judged too coarse or based on the wrong relevant difference. This does not make earlier comparison useless, but it means that comparative claims should be indexed to the scheme under which they were made. Future work should therefore treat comparative-scheme selection as an explicit epistemic problem rather than an invisible preprocessing step.

10.4 Historical Revisability and Epistemic Progress

The claim that recognizability is historically revisable does not entail that later interpretations are automatically better. History can produce new measurements, new concepts, and more discriminating question spaces. It can also produce forgotten evidence, institutional bias, politically stabilized classifications, or technically sophisticated forms of misrecognition. A historical epistemology of suffering must therefore distinguish revisability from progress.

The state vector introduced in Section 9,

$$Z_t = (\mathcal{E}_t, C_t, Q_t, I_t, \mathcal{P}_t),$$

may change in many directions. An increase in evidential quantity can coincide with a poorer conceptual partition. A new classification can improve prediction while narrowing attention to dimensions that are institutionally convenient rather than experientially relevant. A later question space can be richer in one respect and more exclusionary in another. Consequently,

$$t_2 > t_1$$

does not by itself license

$$\text{Adequacy}(Z_{t_2}) > \text{Adequacy}(Z_{t_1}).$$

Any claim of epistemic improvement needs an argument about what has improved and relative to which evidential or interpretive task.

This limitation matters for retrospective judgment. A later framework may render an earlier case newly recognizable, but later recognizability does not imply that earlier observers possessed the same concepts, evidence, or interpretive options. The present paper is concerned with the history of recognition, not with retrospective assignment of responsibility. Questions about whether an earlier observer should have known better, which evidence was reasonably available, or whether ignorance was culpable belong to a separate normative analysis.

Historical openness also places a limit on present confidence. The paper can recommend that recognitional claims be stated with their evidential and interpretive conditions, but it cannot predict which future concepts will reorganize today's cases. The possibility of future re-recognition should therefore encourage explicit provenance and conceptual humility. It should not be converted into the claim that all present classifications are equally unstable or that judgment should be indefinitely suspended.

10.5 Normative Scope of the Epistemic Account

The most important boundary of the paper concerns normativity. Suffering is ethically and practically important in many domains, but an epistemology of suffering recognition does not by itself establish what should be done after suffering is recognized. The following implications remain invalid without additional normative premises:

$$\hat{S}_B \not\Rightarrow \text{moral patienthood}_B,$$

$$\hat{S}_B \not\Rightarrow \text{duty to aid}_H,$$

and

$$\text{duty to aid}_H \not\Rightarrow \text{authority to intervene}_H.$$

These gaps are not technical omissions that can be filled by accumulating more observations. They concern the normative bridges connecting recognized condition, moral relevance, obligation, permission, authority, and intervention.

The same caution applies to relational harm. Demonstrating that a condition narrows a subject's future relations or produces persistent functional impairment may strengthen the case that something normatively significant is occurring, but the present paper does not establish a universal rule under which every reduction of generativity counts as injustice. Constraint can be protective, chosen, temporary, developmentally necessary, or part of a subject's own valued practice. A later ethical account must therefore investigate how relational evidence enters a normative system rather than treating the evidence itself as a complete norm.

This boundary also protects the paper from an inverse error. If phenomenal certainty is unavailable, it does not follow that no ethical question can arise. Ethical and humanitarian theories may adopt precautionary, rights-based, relational, capability-based, care-based, or other normative bridges under uncertainty. Evaluating those bridges belongs to the subsequent paper on humanitarian aid under epistemic uncertainty. The present inquiry ends one step earlier: with the structure and limits of recognition.

10.6 Research Extensions

Several extensions follow from these limits. First, the framework needs empirical operationalization in domains where longitudinal and intervention-sensitive data are available. Animal pain research is an obvious candidate because it already combines behavioral, physiological, pharmacological, and contextual indicators while confronting species-specific expression and validation problems [41, 38]. Such work could test whether the proposed distinction among manifestational, functional, and relational evidence improves analysis or merely redescribes existing practice.

Second, artificial-agent cases require a more precise account of what counts as persistence, preference, aversion, identity continuity, and state restoration in systems whose architecture differs substantially from biological organisms. Current artificial systems can produce human-like reports without establishing the same relation between report and phenomenal state found in human testimony. Future systems may add persistent memory, self-modification, long-horizon planning, or machine-specific communication channels. The framework should therefore remain open to revision rather than treating present-day language models as the canonical artificial case.

Third, cross-interpretive comparison requires tools for testing robustness across alternative quotients. A useful future analysis would ask whether a candidate adverse-state attribution survives changes in the feature map Φ_Q , temporal granularity, intervention model, or observer background assumptions. Robustness across reasonable schemes would not prove phenomenal identity, but it could strengthen the claim that the observed relational pattern is not an artifact of one arbitrary partition.

Fourth, historical analysis raises an archival problem. If future concepts may make present manifestations newly recognizable, then what traces should be preserved today? The present paper does not develop a normative theory of record preservation, but its argument suggests an epistemic reason to distinguish raw observations from current interpretation and to preserve provenance where feasible. This issue connects suffering recognition to a broader GR research program on recognizability, reporting, archives, and the conditions of future inquiry.

Finally, the paper leaves open the relation between recognizability and action under uncertainty. Recognition can be incomplete while decisions cannot always be postponed. Determining when uncertainty supports observation, temporary protection, non-intervention, further inquiry, or stronger intervention requires normative and institutional arguments beyond the present scope. Those questions motivate the next stage of the series rather than a conclusion of this one.

Taken together, these limits preserve the intended modesty of the framework. Generative relational epistemology can clarify how suffering becomes visible, how recognition depends on interpretation and history, how heterogeneous cases may be compared, and where those comparisons remain fragile. It does not eliminate phenomenal opacity, choose the uniquely correct quotient, guarantee historical progress, or derive humanitarian authority from recognition alone. Those unresolved points are not peripheral qualifications. They define the boundary between the epistemic problem addressed here and the ethical, legal, and governance problems that follow from it.

11 Conclusion

This paper has treated suffering first as a problem of recognition. The motivating difficulty is simple to state but difficult to resolve: another subject's suffering is rarely given to an observer as an unmediated fact. Human observers encounter manifestations, reports, bodily changes, functional disruptions, behavioral trajectories, and relational effects, and they interpret those materials through historically developed concepts and background knowledge. The epistemic problem becomes especially visible when the subject differs from the observer in embodiment, communication, cognition, or mode of existence, but the problem is already present within ordinary human recognition. Shared embodiment, language, clinical practice, and social familiarity make human suffering comparatively readable; they do not remove the distinction between another's experience and the observer's recognition of it.

The Generative Relational approach developed here therefore began by suspending rather than resolving the ontological status of phenomenal experience. This suspension was not intended to deny phenomenal reality or reduce suffering to behavior. It served a narrower purpose: to prevent the analysis from requiring a completed theory of consciousness before it could investigate how suffering becomes recognizable. On that basis, the paper distinguished phenomenal suffering, manifested suffering, functional adverse states, and relational or trajectory-level harm. These dimensions can interact, support one another evidentially, and become mutually informative over time, but they remain analytically non-equivalent. Functional impairment is not an objective substitute for pain, visible distress is not transparent access to experience, and relational harm is not a redefinition of phenomenality.

This separation made it possible to examine recognition without privileging one manifestation channel. Human-readable suffering can be epistemically useful because humans possess dense interpretive resources for human expression, yet ease of recognition is not a measure of the reality or severity of an adverse state. Cross-species comparison makes this limitation visible. Animals can express adverse conditions through species-specific postures, facial configurations, activity patterns, protective behavior, or changes in social interaction. A hypothetical alien case extends the problem by removing familiar morphology and communication altogether. Artificial systems create a different asymmetry: they may produce highly human-readable linguistic manifestations while the relation between those manifestations and phenomenal experience remains unsettled. Across these cases, recognizability and phenomenal status cannot be ordered by surface resemblance alone.

The analysis of interpretive inversion sharpened this point. An observer may not merely fail to recognize vulnerability; the observer may classify its manifestation as the opposite kind of state. Agitation, aggression, withdrawal, silence, or unfamiliar physiological change can be placed into categories such as threat, malfunction, defiance, or harmless difference even when later evidence supports a distress-related interpretation.

The relevant lesson is not that aggression is secretly suffering, nor that every unusual behavior should be reclassified as pain. It is that classification is an epistemic operation performed by an interpreter, and the adequacy of that operation cannot be inferred from the salience or familiarity of the manifestation itself.

Semantic opacity does not end the inquiry. When the observer cannot understand a subject's language or internal meaning, portions of the relational dynamics may remain accessible. Persistence, avoidance, willingness to incur costs, intervention-sensitive change, longer-horizon reorganization, and contextual dependence can contribute to a reconstruction of an adverse relational pattern. Such reconstruction remains fallible and model-dependent. It can be confounded by learning, novelty, architecture, environmental constraint, or mechanisms unrelated to phenomenal suffering. Nevertheless, it provides an intermediate epistemic layer between two unsatisfactory extremes: treating unfamiliar subjects as wholly unknowable, and projecting human categories directly onto them. The result is a constrained relational hypothesis rather than a transparent reading of first-person experience.

Cross-interpretive comparison then requires an account of what counts as relevant similarity. Complete structural identity is generally unavailable in historically changing relational systems, while surface resemblance is too weak and anthropocentrically biased. The paper therefore used quotient-based comparison to make the comparative scheme explicit. Two relational histories may be treated as equivalent only relative to selected features and a specified inferential purpose. A manifestation-level quotient, a functional quotient, and a relational quotient can classify the same pair of cases differently. Equivalence at one level does not propagate automatically to another. The governing methodological question is therefore not whether two cases are simply the same, but what differences have been set aside and what structure has been preserved for the present inquiry.

This comparative problem is also historical. Recognizability changes when evidence changes, but it can also change when concepts, classification schemes, interpretive practices, and the available space of questions change. The historical development of pain recognition therefore cannot be represented only as the accumulation of more observations around a fixed conceptual target. A later community may possess categories that allow previously discounted manifestations to become intelligible, or it may formulate questions that an earlier community could not yet ask. Conversely, later frameworks can introduce new blind spots. Historical revisability is therefore not equivalent to inevitable epistemic progress. It indicates that recognitional claims should be understood together with the evidential and conceptual conditions under which they became possible.

Taken together, these arguments support a conception of suffering recognition as a temporally situated relation among subject, manifestation, observer, evidence, and interpretive system. The observer does not encounter an already labelled object called suffering and then merely report it. Recognition is generated through a sequence in which manifestations become evidence, evidence is organized through concepts and comparative schemes, and resulting interpretations can alter subsequent observation and inquiry. This generative process is constrained by the world: interpretations can fail, interventions can reveal unexpected causal structure, and trajectories can resist an observer's preferred classification. Yet constraint does not eliminate mediation. Recognition remains relationally achieved rather than epistemically transparent.

The contribution of this paper is consequently methodological and epistemic. It does not provide a universal suffering detector, a scalar measure applicable across all subjects, or a criterion that settles phenomenal consciousness through relational evidence. It does not identify one uniquely correct quotient for cross-species comparison. It also

does not imply that later interpretations are automatically superior to earlier ones. Instead, it offers a vocabulary for separating dimensions that are often collapsed, for making comparative assumptions explicit, and for tracing how recognizability itself develops through changing histories of evidence and interpretation.

The same restraint marks the boundary between this paper and the ethical questions that motivated it. Recognizing a candidate adverse state does not by itself establish moral patienthood, a duty to provide aid, permission to intervene, or authority to transform another subject's relational world. Those transitions require additional normative premises. An epistemology of suffering can clarify what is being recognized, how confidently, through which interpretive resources, and with which unresolved uncertainties; it cannot by itself supply the normative bridge from recognition to action. The next stage of the broader research program therefore begins where the present paper ends: with the question of how recognized or suspected suffering becomes normatively salient, how obligations of aid are generated, and under what conditions concern for another subject can justify intervention.

The broader implication is modest but consequential. Ethical and humanitarian reasoning often begins from a sentence such as "the other is suffering." Before that sentence can bear normative weight, however, there is already a history of manifestation, observation, classification, comparison, and interpretation. Making that history visible does not dissolve suffering into interpretation. It clarifies the epistemic conditions under which suffering becomes recognizable and the points at which recognition can fail. For heterogeneous subjects in particular, such clarification is a necessary preliminary to any responsible account of what ought to follow.

Responsible Use and Rights Reservation

This section separates requested scholarly conduct from the legal permissions stated on the following page. It records an ethical request for responsible use and then defines the narrower scope of retained legal rights.

The author encourages good-faith discussion, criticism, independent inquiry, and responsible use of the material in this work. Separately from the licence's terms, the author asks users to consider foreseeable harms when adapting or applying the proposed framework. This ethical request leaves the licence's permissions and legally authorized uses unchanged.

The author retains the rights preserved under CC BY-NC 4.0 and may pursue remedies to which the author is legally entitled for breach of the licence or violation of the author's independently applicable rights. Reuse remains independent from authorial endorsement. Third-party rights require authorization from their respective holders where applicable. Copyright exceptions and limitations, including applicable forms of fair use or fair dealing, remain fully available.

Notices

This page consolidates the manuscript’s publication status, licence, development disclosure, research-programme relation, and suggested citation.

Status. This working draft records an evolving stage of the author’s position and is circulated for discussion. Definitions, section structure, formal distinctions, and terminology remain subject to revision. Literature verification, conceptual history, comparison with existing theories of pain and suffering, cross-species recognition research, AI and animal-welfare scholarship, and specialist review remain future research stages.

Licence. Except where otherwise indicated, copyright 2026 Wanhong Huang. This work is made available under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). Subject to its terms, the licence permits sharing and adaptation for noncommercial purposes with appropriate attribution, a link to the licence, an indication of changes, and attribution that preserves the licensor’s independence from the reuse. Reuse is governed solely by that licence; the responsible-use request on the preceding page remains separate from its terms. The licence deed and legal-code link are available at <https://creativecommons.org/licenses/by-nc/4.0/>. The licence governs in case of conflict with this summary. Third-party material remains subject to the rights held by its respective rights holders.

Statement on the use of language models. The exploratory discussions and preparation of this paper involved OpenAI’s ChatGPT. ChatGPT supported exploratory dialogue, formal reconstruction, source discovery followed by website verification, argumentative criticism, and drafting in $\text{\LaTeX}$. The author selected the research questions, directed and approved the theoretical commitments and epistemic status of the claims, and bears sole responsibility for the manuscript, including its definitions, formal constructions, taxonomy, arguments, conclusions, and errors. Authorship credit remains with the human author.

References

- [1] Clare S. Allely. Pain sensitivity and observer perception of pain in individuals with autistic spectrum disorder. *The Scientific World Journal*, page 916178, 2013.
- [2] K. J. S. Anand and P. R. Hickey. Pain and its effects in the human neonate and fetus. *The New England Journal of Medicine*, 317(21):1321–1329, 1987.
- [3] Anita Avramides. Other minds. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2023 edition, 2023.
- [4] Jonathan Birch, Charlotte Burn, Alexandra Schnell, Heather Browning, and Andrew Crump. Review of the evidence of sentience in cephalopod molluscs and decapod crustaceans. Technical report, London School of Economics and Political Science, London, 2021.
- [5] Christophe Boesch. What makes us human (*Homo sapiens*)? the challenge of cognitive cross-species comparison. *Journal of Comparative Psychology*, 121(3):227–240, 2007.

- [6] S. Brown and Jonathan Birch. When and why are motivational trade-offs evidence of sentience? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 380(1939), 2025.
- [7] Patrick Butlin, Robert Long, Eric Elmoznino, Yoshua Bengio, Jonathan Birch, Axel Constant, George Deane, Stephen M. Fleming, Chris Frith, Xu Ji, Ryota Kanai, Colin Klein, Grace Lindsay, Matthias Michel, Liad Mudrik, Megan A. K. Peters, Eric Schwitzgebel, Jonathan Simon, and Rufin VanRullen. Consciousness in artificial intelligence: Insights from the science of consciousness. arXiv preprint arXiv:2308.08708, 2023.
- [8] Eric J. Cassell. The nature of suffering and the goals of medicine. *The New England Journal of Medicine*, 306(11):639–645, 1982.
- [9] Lorraine Daston and Peter Galison. *Objectivity*. Zone Books, New York, 2007.
- [10] Mark E. Epstein, Ilona Rodan, Greg Griffenhagen, Jamie Kadrlik, Michael C. Petty, Sheilah A. Robertson, and Wendy Simpson. 2015 AAHA/AAFP pain management guidelines for dogs and cats. *Journal of the American Animal Hospital Association*, 51(2):67–84, 2015.
- [11] Marina C. Evangelista, Beatriz P. Monteiro, and Paulo V. Steagall. Measurement properties of grimace scales for pain assessment in nonhuman mammals: A systematic review. *Pain*, 163(6):e697–e714, 2022.
- [12] G. Allen Finley, Ólöf Kristjánsdóttir, and Paula A. Forgeron. Cultural influences on the assessment of children’s pain. *Pain Research and Management*, 14(1):33–37, 2009.
- [13] Carola Fischer-Tenhagen, Jennifer Meier, and Alina Pohl. “Do Not Look at Me Like That”: Is the facial expression score reliable and accurate to evaluate pain in large domestic animals? a systematic review. *Frontiers in Veterinary Science*, 9:1002681, 2022.
- [14] Ludwik Fleck. *Genesis and Development of a Scientific Fact*. University of Chicago Press, Chicago, 1981.
- [15] Elisabeth Flo, Christine Gulla, and Bettina S. Husebø. Effective pain management in patients with dementia: Benefits beyond pain? *Drugs & Aging*, 31(12):863–871, 2014.
- [16] Roman Frigg and James Nguyen. *Modelling Nature: An Opinionated Introduction to Scientific Representation*, volume 427 of *Synthese Library*. Springer, Cham, 2020.
- [17] Kazuhiro Goto. How can anthropomorphism be incorporated into the comparative cognition studies? *Japanese Journal of Animal Psychology*, 62(1):49–57, 2012.
- [18] Ian Hacking. *Historical Ontology*. Harvard University Press, Cambridge, MA, 2002.
- [19] Thomas Hadjistavropoulos and Kenneth D. Craig. A theoretical framework for understanding self-report and observational measures of pain: A communications model. *Behaviour Research and Therapy*, 40(5):551–570, 2002.
- [20] Thomas Hadjistavropoulos, Kenneth D. Craig, Steve Duck, Annmarie Cano, Liesbet Goubert, Philip L. Jackson, Jeffrey S. Mogil, Pierre Rainville, Michael J. L. Sullivan, Amanda C. de C. Williams, Tine Vervoort, and Theresa Dever Fitzgerald. A biopsychosocial formulation of pain communication. *Psychological Bulletin*, 137(6):910–939, 2011.

- [21] I. Hernández-Avalos, D. Mota-Rojas, J. E. Mendoza-Flores, A. Casas-Alvarado, K. Flores-Padilla, A. E. Miranda-Cortes, F. Torres-Bernal, J. Gómez-Prado, and P. Mora-Medina. Nociceptive pain and anxiety in equines: Physiological and behavioral alterations. *Veterinary World*, 14(11):2984–2995, 2021.
- [22] Wanhong Huang. The relational mode of being of concepts: Towards a generative relational epistemology. Zenodo preprint, 2026.
- [23] International Committee of the Red Cross. Our fundamental principles. International Committee of the Red Cross. Accessed 15 September 2026.
- [24] Celeste C. Johnston. Neonatal pain: A journey spanning three decades. *Paediatric and Neonatal Pain*, 2(2):33–39, 2020.
- [25] Nicholas V. Karayannis, John A. Sturgeon, Ming Chih-Kao, Corinne Cooley, and Sean C. Mackey. Pain interference and physical function demonstrate poor longitudinal association in people living with pain: A PROMIS investigation. *Pain*, 158(6):1063–1068, 2017.
- [26] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 4 edition, 2012.
- [27] Julia Kumar, Cheryl Hartzell, Elana Abelson, Kristen Jastrowski Mano, and Vidya Chidambaran. Prevalence, expression, assessment, mechanisms, and management of pain in autistic children: A scoping review. *A&A Practice*, 20(4):e02178, 2026.
- [28] Robert Long, Jeff Sebo, Patrick Butlin, Kathleen Finlinson, Kyle Fish, Jacqueline Harding, Jacob Pfau, Toni Sims, Jonathan Birch, and David Chalmers. Taking ai welfare seriously. arXiv preprint arXiv:2411.00986, 2024.
- [29] Daniel S. Mills, Fergus M. Coutts, and Kevin J. McPeake. Behavior problems associated with pain and paresthesia. *Veterinary Clinics of North America: Small Animal Practice*, 54(1):55–69, 2024.
- [30] Daniel S. Mills, Isabelle Demontigny-Bédard, Margaret Gruen, Mary P. Klinck, Kevin J. McPeake, Ana M. Barcelos, Lorraine Hewison, Hanne Van Haevermaet, Sagi Denenberg, Helene Hauser, Charlotte Koch, and Kelly Ballantyne. Pain and problem behavior in cats and dogs. *Animals*, 10(2):318, 2020.
- [31] Jeffrey S. Mogil, Daniel S. J. Pang, Gabrielle Guanaes Silva Dutra, and Christine T. Chambers. The development and use of facial grimace scales for pain measurement in animals. *Neuroscience and Biobehavioral Reviews*, 116:480–493, 2020.
- [32] Thomas Nagel. What is it like to be a bat? *The Philosophical Review*, 83(4):435–450, 1974.
- [33] James Nguyen. Scientific representation and theoretical equivalence. *Philosophy of Science*, 84(5):982–995, 2017.
- [34] Kenneth M. Prkachin. Facial pain expression. *Pain Management*, 1(4):367–376, 2011.
- [35] Srinivasa N. Raja, Daniel B. Carr, Milton Cohen, Nanna B. Finnerup, Herta Flor, Stephen Gibson, Francis J. Keefe, Jeffrey S. Mogil, Matthias Ringkamp, Kathleen A. Sluka, Xue-Jun Song, Bonnie Stevens, Mark D. Sullivan, Perri R. Tutelman, Takahiro Ushida, and Kyle Vader. The revised international association for the study of pain definition of pain: Concepts, challenges, and compromises. *Pain*, 161(9):1976–1982, 2020.

- [36] Elissa N. Rodkey and Rebecca Pillai Riddell. The infancy of infant pain research: The experimental origins of infant pain denial. *The Journal of Pain*, 14(4):338–350, 2013.
- [37] Álvaro Sabater-Gárriz, Jesús Molina-Mula, Pedro Montoya, and Inmaculada Riquelme. Pain assessment tools in adults with communication disorders: Systematic review and meta-analysis. *BMC Neurology*, 24(1):66, 2024.
- [38] Katelyn E. Sadler, Jeffrey S. Mogil, and Cheryl L. Stucky. Innovations and advances in modelling and measuring pain in animals. *Nature Reviews Neuroscience*, 23:70–85, 2022.
- [39] Christelle Salabura, Arnaud Sourty, Flora Bat-Pitault, Kirsty Regnery, Sandrine Mayen, and Sébastien Colson. Pain assessment for children and adolescents with autism spectrum disorders (ASD): A systematic review. *L'Encéphale*, 51(1):87–94, 2025.
- [40] Tianyue Shi, Ye Xu, Qianping Li, Ling Zhu, Hongfei Jia, Kai Qian, Siwen Shi, Xianwen Li, Yueheng Yin, and Yaping Ding. Association between pain and behavioral and psychological symptoms of dementia (BPSD) in older adults with dementia: A systematic review and meta-analysis. *BMC Geriatrics*, 25(1):100, 2025.
- [41] Lynne U. Sneddon, Robert W. Elwood, Shelley A. Adamo, and Matthew C. Leach. Defining and assessing animal pain. *Animal Behaviour*, 97:201–212, 2014.
- [42] Liv Inger Strand, Kim Fredrik Gundrosen, Regina Küfner Lein, Marjan Laekeman, Frank Lobbezoo, Ruth Defrin, and Bettina S. Husebø. Body movements as pain indicators in older people with cognitive impairment: A systematic review. *European Journal of Pain*, 23(4):669–685, 2019.
- [43] Kevin M. Summers and E. Paige Lloyd. Pain expression and decoding as intergroup phenomena. *Affective Science*, 2026. Online ahead of print.
- [44] Tyler Tate. What really is the nature of suffering? three problems with eric cassell's concept of distress. *Bioethics*, 2020.
- [45] John Worrall. Structural realism: The best of both worlds? *Dialectica*, 43(1–2):99–124, 1989.