UNITED STATES PROVISIONAL PATENT APPLICATION

CONTEXTUAL COHERENCE FIELDS WITH DUAL-PROCESS COGNITIVE ARCHITECTURE AND MANIFOLD-CONSTRAINED COHERENCE MIXING: SYSTEM AND METHOD FOR CONTEXT-KEYED RELATIONAL COHERENCE ACCUMULATION, DOUBLY STOCHASTIC CROSS-CONTEXT TRANSFER, AND BIDIRECTIONAL REFLEXIVE-DELIBERATIVE PROCESSING IN AUTONOMOUS SOCIAL ROBOTS

Inventor: Colm Byrne Flout Labs (Flout Ltd) Cave, Clarinbridge, Co. Galway, Ireland

Filing Date: February 23, 2026

FIELD OF THE INVENTION

[0001] The present invention relates generally to the field of autonomous social robotics and human-robot interaction (HRI). More specifically, the invention concerns systems and methods for managing the behavioral expressiveness of an autonomous robot through context-keyed coherence accumulators that independently track accumulated relational familiarity across distinct environmental and social contexts, with optional manifold-constrained cross-context coherence mixing that permits bounded transfer of relational familiarity between related contexts while preserving total coherence energy, combined with a bidirectional dual-process cognitive architecture wherein a reflexive processing unit and a deliberative processing unit mutually modulate each other's outputs, enabling emergent personality expression, learned inhibition, behavioral compilation, and offline consolidation without static personality parameterization.

BACKGROUND OF THE INVENTION

[0002] The integration of autonomous robots into domestic and social environments has created a demand for robots capable of sustained, naturalistic interaction with human occupants. A central challenge in this domain is how a robot should modulate its behavioral expressiveness over time and across different social situations to produce interactions that humans perceive as appropriate, trustworthy, and increasingly familiar.

[0003] Existing approaches to managing robot behavioral expressiveness fall into three broad categories, each with significant limitations.

[0004] The first category comprises static personality parameterization systems. In these systems, a robot's behavioral tendencies are defined by a fixed set of trait values, commonly derived from psychological frameworks such as the Big Five personality model (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). For example, a robot may be configured with an extraversion parameter set to 0.3, producing uniformly reserved behavior regardless of context, history, or accumulated interaction. These systems are incapable of differentiating between contexts: a robot configured for low extraversion will behave identically toward a familiar household member and a complete stranger, failing to capture the context-specific nature of relational comfort. The dominance of static personality parameterization in the field is evidenced by representative works including: Moshkina and Arkin (2005), who proposed a trait-based affective architecture for social robots using Big Five personality parameters to modulate behavior generation, with traits remaining fixed throughout the robot's operational life; Lee et al. (2006), who demonstrated that matching a robot's Big Five personality profile to a user's personality improved perceived social presence, but implemented personality as a static configuration selected at deployment time; Tapus and Mataric (2008), who proposed personality-based adaptive robot behavior for post-stroke rehabilitation, wherein the robot's extraversion-introversion axis was pre-configured and remained constant regardless of accumulated interaction history; and Joosse et al. (2013), who mapped Big Five personality traits to specific robot behaviors (e.g., extraversion to approach distance, agreeableness to compliance) using fixed parameter tables. In each case, the prevailing approach in the field treats personality as a design-time configuration rather than an emergent property of accumulated relational experience, and no mechanism exists for context-specific behavioral differentiation based on earned familiarity.

[0005] The second category comprises global emotional state systems. In these systems, the robot maintains a small set of scalar values representing its current emotional or arousal state, typically computed as a function of recent sensor input. A representative implementation

computes a single coherence value as an exponentially-weighted moving average of tension instability derived from sensor variance. While responsive to moment-to-moment environmental conditions, global emotional state suffers from three critical deficiencies: (a) a transfer problem, wherein comfort accumulated through interaction with one person or in one context inappropriately transfers to a different person or context; (b) context collapse, wherein temporally or situationally distinct interaction contexts are conflated into a single state; and (c) regression amnesia, wherein a single negative event resets all accumulated relational progress regardless of the depth of prior positive interaction history.

[0006] The third category comprises multi-stream residual architectures in deep neural networks, exemplified by Hyper-Connections (Zhu et al., 2024) and Manifold-Constrained Hyper-Connections (Xie et al., 2026). These architectures expand the width of information streams flowing through a network and introduce learnable mixing matrices to govern cross-stream information exchange. While demonstrating significant performance gains in language model training, these architectures have not been applied to the domain of social robot behavioral management, and the mathematical properties of manifold-constrained mixing (norm preservation, compositional closure, and Birkhoff polytope geometry) have not been recognized as applicable to the problem of context-specific relational trust accumulation in embodied autonomous agents.

[0007] A further limitation of existing approaches is the absence of a principled cognitive architecture governing the interaction between reflexive and deliberative processing in social robots. Existing systems typically employ either a purely reactive architecture (stimulus-response with no deliberation) or a purely deliberative architecture (all behavior computed from models and plans). Neither captures the richness of biological cognitive systems, where behavior emerges from the continuous, bidirectional interaction between fast reflexive processes and slower deliberative processes, with each system capable of modulating, inhibiting, or overriding the other based on accumulated experience and current context.

[0008] Neither static personality parameters, global emotional state systems, unconstrained multi-stream architectures, nor existing cognitive architectures are capable of producing the phenomenon described herein as "earned fluency": a gradual, context-specific accumulation of behavioral expressiveness that must be independently developed in each distinct relational context the robot encounters, that cannot be transferred across contexts without manifold-constrained bounds, that is resilient to isolated negative events proportional to the depth of accumulated positive history, and that emerges from the bidirectional interaction between reflexive and deliberative processing subsystems.

[0009] What is needed is a system that replaces global coherence with a map of independently earned, independently persistent, and independently losable coherence accumulators, keyed to detectable environmental and social context features, optionally connected through manifold-constrained mixing that permits bounded cross-context transfer while preserving total coherence energy, gated by an architectural constraint that prevents accumulated familiarity from being expressed in unfamiliar contexts, and managed by a dual-process cognitive architecture wherein reflexive and deliberative subsystems mutually modulate each other's outputs.

SUMMARY OF THE INVENTION

[0010] The present invention provides a system and method for managing the behavioral expressiveness of an autonomous social robot through Contextual Coherence Fields (CCF) operating within a dual-process cognitive architecture. The system comprises six primary components operating in coordination.

[0011] First, a context detection subsystem that constructs composite context keys from multiple sensor signals, creating a situation fingerprint representing the robot's current environmental and social context.

[0012] Second, a coherence accumulation subsystem comprising a plurality of context-keyed coherence accumulators, each independently tracking accumulated relational familiarity for a specific context, with accumulation dynamics modulated by configurable personality parameters and protected by an interaction-count-proportional decay floor that prevents catastrophic loss of deeply accumulated relational progress.

[0013] Third, a behavioral gating subsystem that computes an effective coherence value as the minimum of an instantaneous coherence (derived from current sensor stability) and the accumulated context coherence for the current context, enforcing the architectural invariant that the robot can never express more behavioral richness than its accumulated familiarity with the present context permits.

[0014] Fourth, a manifold-constrained coherence mixing subsystem that, in a further embodiment, organizes the plurality of coherence accumulators into a coherence vector and applies a doubly stochastic mixing matrix to govern bounded cross-context coherence transfer, the mixing matrix being constrained to the Birkhoff polytope via Sinkhorn-Knopp projection such that (a) the spectral norm of the mixing matrix is bounded by 1, preventing coherence amplification; (b) the set of doubly stochastic matrices is closed under composition, ensuring long-term behavioral stability; and (c) the mixing matrix admits decomposition as a convex

combination of permutation matrices, providing geometric interpretability of cross-context transfer patterns. The mixing matrix comprises a static component representing baseline cross-context relationships and a dynamic component modulated by current sensor input, with the dynamic component gated by a factor initialized near zero such that the robot architecturally begins with minimal cross-context transfer and earns increased transfer capacity through accumulated interaction.

[0015] Fifth, a context boundary discovery subsystem that constructs a relational graph from accumulated interaction episodes, computes trajectory embeddings representing behavioral patterns within each context, and applies graph min-cut algorithms to automatically discover natural context boundaries. In the manifold-constrained embodiment, context boundaries are additionally derivable from the structure of the doubly stochastic mixing matrix, wherein near-zero entries identify boundaries between coherence domains and non-zero entries quantify the degree of cross-context behavioral relationship.

[0016] Sixth, a dual-process cognitive architecture comprising a reflexive processing unit and a deliberative processing unit that operate in bidirectional modulation. The reflexive processing unit executes real-time sensor processing, context detection, coherence gating, and behavioral output at each processing tick. The deliberative processing unit performs persistent memory management, trajectory embedding computation, context boundary discovery, mixing matrix parameter optimization, learned inhibition map generation, behavioral sequence compilation, and offline consolidation of accumulated experience. The two units mutually modulate each other's outputs: the deliberative unit pre-loads context-specific stimulus suppression maps, compiled behavioral sequences, and updated mixing matrix parameters to the reflexive unit, while the reflexive unit generates interaction episodes, conflict signals, and mixing matrix divergence signals that inform the deliberative unit's processing.

[0017] In a further aspect, the effective coherence value and a concurrent tension value define a two-dimensional behavioral phase space partitioned into four behavioral quadrants with hysteresis-based boundary transitions, each quadrant producing a distinct behavioral profile ranging from cautious observation to full expressive fluency.

[0018] In a further aspect, the manifold-constrained coherence mixing subsystem enables a multi-dimensional behavioral phase space wherein the full coherence vector, rather than a single scalar effective coherence, modulates behavioral output, permitting richer behavioral differentiation based on the robot's simultaneous relationship to multiple context domains.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] FIG. 1 is a block diagram illustrating the overall system architecture of the Contextual Coherence Fields system with dual-process cognitive architecture and manifold-constrained coherence mixing, showing the relationship between the reflexive processing unit, the deliberative processing unit, the coherence mixing subsystem, and the bidirectional modulation pathways between them.

[0020] FIG. 2 is a diagram illustrating the two-dimensional behavioral phase space defined by context coherence and tension axes, showing four behavioral quadrants: Shy Observer, Startled Retreat, Quietly Beloved, and Protective Guardian, with hysteresis boundaries.

[0021] FIG. 3 is a flow diagram illustrating the per-tick processing cycle of the reflexive processing unit, from sensor input through context key construction, accumulator lookup, optional manifold-constrained mixing, effective coherence computation, learned inhibition application, and behavioral output modulation.

[0022] FIG. 4 is a diagram illustrating the construction of a relational graph from accumulated interaction episodes and the application of graph min-cut to discover coherence group boundaries, with correspondence to the structure of the doubly stochastic mixing matrix in the manifold-constrained embodiment.

[0023] FIG. 5 is a graph illustrating coherence accumulator dynamics over time, showing asymptotic growth, personality-modulated deltas, interaction-count-proportional decay floor, and recovery from negative events.

[0024] FIG. 6 is a flow diagram illustrating the bidirectional modulation pathways of the dual-process cognitive architecture, showing the downward pathway (deliberative to reflexive: learned inhibition maps, compiled behavioral sequences, updated coherence accumulators, mixing matrix parameters) and the upward pathway (reflexive to deliberative: interaction episodes, conflict signals, mixing divergence signals, novel context notifications).

[0025] FIG. 7 is a diagram illustrating the habit compilation lifecycle, showing how a behavioral sequence progresses from deliberative processing through repetition thresholding to compiled reflexive execution, with automatic fallback to deliberative processing upon distribution shift detection.

[0026] FIG. 8 is a timeline diagram illustrating a consolidation cycle during an idle period, showing experience replay, relational graph recomputation, coherence group reorganization, mixing matrix optimization, and stimulus suppression map updates.

[0027] FIG. 9 is a state diagram illustrating the classification conflict resolution mechanism, showing divergent assessments by the reflexive and deliberative processing units, the observable behavioral trace (hesitation, pause, resolution), and the feedback pathway that records the conflict outcome for future learning.

[0028] FIG. 10 is a diagram illustrating the manifold-constrained coherence mixing architecture, showing n coherence streams, the doubly stochastic mixing matrix constrained to the Birkhoff polytope, the Sinkhorn-Knopp projection process, and the relationship between mixing matrix structure and context boundary topology.

[0029] FIG. 11 is a diagram illustrating the evolution of the mixing matrix over the operational life of the robot, showing the initial near-identity state (minimal cross-context transfer), intermediate states with learned partial transfer, and the mature state with rich cross-context relationships, with near-zero entries corresponding to context boundaries identified by min-cut analysis.

[Note: Formal drawings to be prepared by patent illustrator prior to non-provisional filing.]

FIG. 3 - Per-Tick Reflexive Processing Cycle

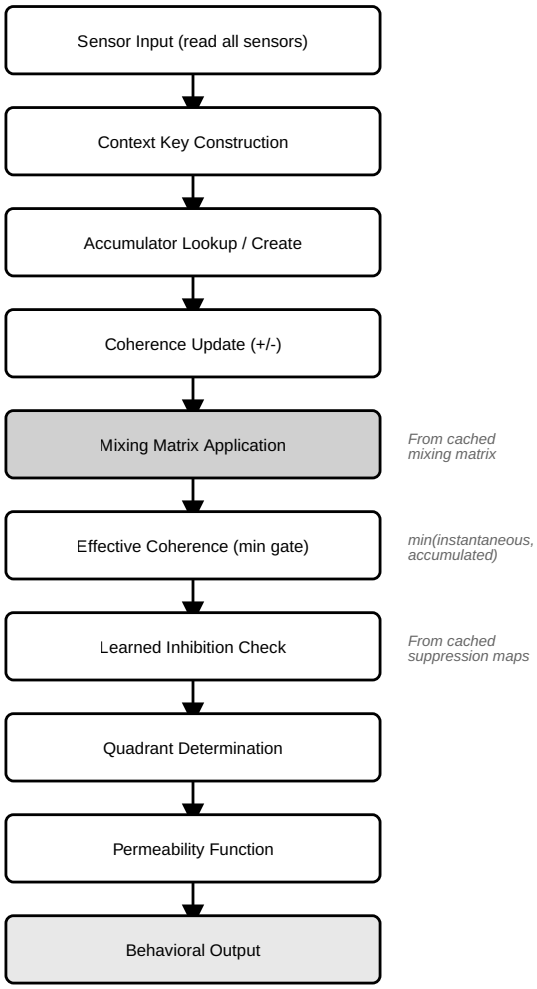

FIG. 3 — Per-tick processing cycle of the reflexive processing unit.

FIG. 5 - Coherence Accumulator Dynamics Over Time

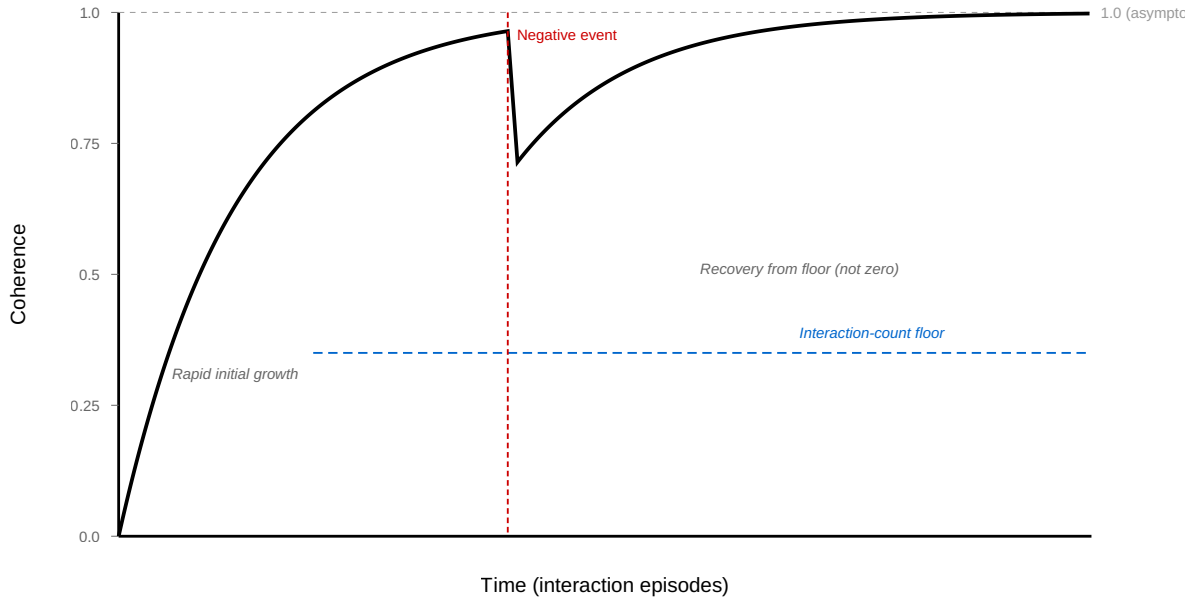

FIG. 5 — Coherence accumulator dynamics: asymptotic growth, earned floor, and recovery from negative events.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

PART I: CONTEXTUAL COHERENCE FIELDS

1. System Overview

[0030] Referring now to FIG. 1, the Contextual Coherence Fields (CCF) system 100 comprises six primary subsystems operating on an autonomous robot platform: a context detection subsystem 110, a coherence accumulation subsystem 120, a behavioral gating subsystem 130, a manifold-constrained coherence mixing subsystem 135, a context boundary discovery subsystem 140, and a dual-process cognitive architecture 200 comprising a reflexive processing unit 210 and a deliberative processing unit 220. The system receives input from a sensor array 105 and produces output through a behavioral output interface 150 that controls motor actuators, visual indicators (LEDs), and audio output.

FIG. 1 - System Architecture

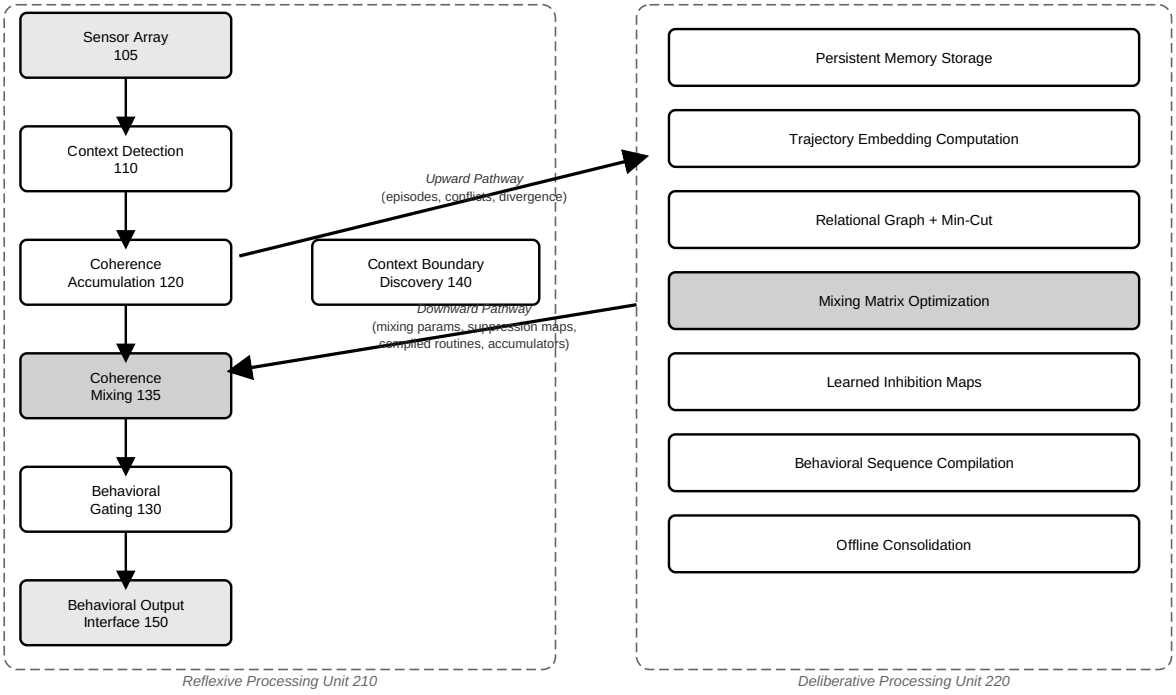

FIG. 1 — Overall system architecture of the Contextual Coherence Fields system with dual-process cognitive architecture.

[0031] In a preferred embodiment, the reflexive processing unit 210 implements the context detection subsystem 110, coherence accumulation subsystem 120, behavioral gating subsystem 130, and a cached instance of the manifold-constrained coherence mixing subsystem 135 as deterministic, real-time components operating without dynamic memory allocation (suitable for no_std embedded environments). The deliberative processing unit 220 implements the context boundary discovery subsystem 140, the mixing matrix parameter optimization, and the bidirectional modulation functions described in Part II of this specification, with access to persistent storage and computational resources sufficient for graph algorithms, mixing matrix training, and experience replay.

2. Context Detection Subsystem

[0032] The context detection subsystem 110 constructs a composite context key from multiple sensor signals at each processing tick. In a preferred embodiment, the sensor array 105 includes an ambient light sensor, an ambient sound level sensor, an ultrasonic distance sensor, an inertial measurement unit (accelerometer and gyroscope), and a system clock.

[0033] Each sensor signal is quantized into a discrete set of feature levels. In a preferred embodiment: the ambient light sensor signal is quantized into brightness bands (e.g., dark, dim, bright) based on configurable thresholds; the ambient sound level is quantized into loudness bands (e.g., quiet, moderate, loud); the ultrasonic distance sensor signal is processed over a sliding window to determine a presence signature (e.g., approaching, static, retreating, absent); the accelerometer signal is classified into motion contexts (e.g., self-moving, being-moved, stationary); the gyroscope signal is classified into orientation contexts (e.g., upright, tilted, being-handled); and the system clock is bucketed into temporal periods (e.g., morning, midday, evening, night).

[0034] The composite context key is formed by concatenation of the quantized feature levels into a single string or hash value. For example, a context key might take the form: "bright:quiet:approaching:stationary:upright:morning". This composite key serves as a situation fingerprint that characterizes the robot's current environmental and social context without requiring identification of specific individuals.

[0035] It is a deliberate and architecturally significant feature of the invention that context keys fingerprint situations rather than individuals. Two different persons approaching the robot under identical environmental conditions will produce the same context key. This design is not a limitation arising from sensor poverty but a principled architectural choice with three consequences.

[0035a] First, the system does not make claims about relational specificity that exceed its actual perceptual resolution. A robot that cannot distinguish Person A from Person B does not maintain the fiction that it has separate relationships with each. This honesty principle ensures that the behavioral output of the system is always grounded in what the robot can actually perceive, preventing uncanny misattribution of familiarity.

[0035b] Second, situation fingerprinting captures relational dimensions that individual identification alone cannot. A human behaves differently with the same person in different environmental and temporal contexts: the same friend encountered in a quiet home on a weekend morning and in a noisy office on a weekday afternoon elicits different social behavior. By keying coherence accumulators to the full environmental context rather than to individual identity alone, the system produces context-appropriate behavioral differentiation even within relationships with a single individual, a capability that person-keyed systems lack entirely.

[0035c] Third, and critically for the scope of the invention, the context key vocabulary is designed as an open, extensible schema that scales with sensor capability without architectural modification. On platforms equipped with additional sensors, including but not limited to cameras with facial recognition, microphones with speaker identification, biometric sensors, RFID or NFC readers, or any future sensor modality capable of producing a discrete feature classification, the context key is extended by concatenating additional quantized feature levels to the existing composite key. A camera-equipped platform might produce context keys of the form "bright:quiet:approaching:stationary:upright:morning:PERSON_A", while a resource-constrained platform produces "bright:quiet:approaching:stationary:upright:morning" for the same physical situation. In both cases, the coherence accumulation, behavioral gating, manifold-constrained mixing, and all other subsystems of the invention operate identically on the resulting context key. No component of the system requires modification, reconfiguration, or retraining when the sensor vocabulary changes. The architecture is sensor-agnostic by design: the inventive mechanisms described herein apply to any platform capable of producing at least two quantized sensor features, from a microcontroller with a light sensor and a clock to a humanoid robot with full computer vision and natural language processing.

3. Coherence Accumulation Subsystem

[0036] The coherence accumulation subsystem 120 maintains a map of context keys to coherence accumulators. Each coherence accumulator is an independent data structure comprising: a coherence value (a scalar between 0.0 and 1.0, inclusive), an interaction count (a

non-negative integer tracking the number of interaction episodes recorded in this context), and a last interaction tick (a timestamp recording the most recent interaction in this context).

[0037] At each processing tick, the coherence accumulation subsystem 120 receives the current context key from the context detection subsystem 110 and retrieves or creates the corresponding coherence accumulator. The accumulator is then updated according to the following rules.

[0038] Positive accumulation: When the current sensor state indicates low tension and stable sensor readings (no collision, no startle event), the coherence value is incremented by a positive delta. The positive delta is computed as the product of a base increment rate and a personality-specific recovery speed parameter, further multiplied by the factor (1.0 minus the current coherence value). This multiplicative factor produces asymptotic growth: the coherence value increases rapidly when low and approaches 1.0 with diminishing increments, modeling the natural pattern of trust formation where initial gains are rapid and refinement is gradual.

[0039] Negative accumulation: When the current sensor state indicates a collision, startle event, or high tension, the coherence value is decremented by a negative delta. The negative delta is computed as the product of a base decrement rate and a personality-specific startle sensitivity parameter. A robot with high startle sensitivity loses coherence faster from negative events. A robot with low startle sensitivity is more resilient to perturbation.

[0040] Interaction-count-proportional decay floor: Between interactions, the coherence value decays toward a floor value that is proportional to the accumulated interaction count for that context. The floor value is computed as a function of the interaction count, such that contexts with a deep history of positive interaction maintain a higher minimum coherence even during periods of inactivity or after isolated negative events. This mechanism prevents catastrophic loss of deeply accumulated relational progress: a single negative event cannot erase weeks of positive interaction history. The floor value is bounded above by a maximum floor parameter to prevent accumulators from becoming permanently locked at high values.

[0041] Growth ceiling: The asymptotic growth function (delta multiplied by (1.0 minus current value)) ensures that the coherence value approaches but never reaches 1.0, modeling the phenomenon that perfect relational comfort is asymptotically approached but never fully achieved.

[0042] Personality modulation of dynamics: The accumulator update rules are parameterized by a set of personality parameters, including but not limited to: recovery speed (rate of

positive coherence accumulation), startle sensitivity (rate of negative coherence accumulation), and curiosity drive (which may modulate the rate at which new contexts are created and the initial exploration behavior within low-coherence contexts). Critically, these personality parameters modulate the dynamics of coherence accumulation (how fast it is earned or lost) but do not modify the structural requirement that coherence must be earned. A robot with high curiosity drive earns coherence faster but is not granted coherence without earning it.

4. Behavioral Gating Subsystem

[0043] The behavioral gating subsystem 130 computes an effective coherence value that governs the robot's behavioral output. The effective coherence is computed as the minimum of two values: an instantaneous coherence value and a context coherence value.

[0044] The instantaneous coherence value is computed from the robot's current sensor state, reflecting the stability of the present moment. In a preferred embodiment, the instantaneous coherence is computed as an exponentially-weighted moving average of tension instability, where tension instability is the absolute deviation of the current tension from the tension moving average.

[0045] The context coherence value is the accumulated coherence for the current context key, as maintained by the coherence accumulation subsystem 120. In the manifold-constrained embodiment described in Section 6A, the context coherence value may additionally incorporate bounded contributions from related contexts via the doubly stochastic mixing matrix, as described below.

[0046] The minimum function enforces the central architectural invariant of the system: the robot can never be more behaviorally expressive than the lesser of its current environmental stability and its accumulated relational familiarity with the present context. A calm environment (high instantaneous coherence) combined with an unfamiliar context (low context coherence) produces reserved, cautious behavior. A familiar context (high context coherence) experiencing momentary environmental volatility (low instantaneous coherence) also produces reduced expressiveness. Only when both the current moment is stable and the context has been earned through accumulated positive interaction does the robot express its full behavioral repertoire.

[0046a] The use of a minimum function over instantaneous environmental coherence and accumulated relational coherence as a behavioral gate is, to the inventor's knowledge, without precedent in prior behavioral architectures for autonomous agents, including affective computing systems, social robot controllers, reinforcement learning reward architectures, and

control-theoretic behavioral planners. Prior systems that combine multiple signals into a behavioral output typically employ weighted summation, multiplicative combination, or priority-based arbitration. The minimum function produces a qualitatively distinct behavioral outcome: it creates an asymmetric constraint wherein either factor alone is sufficient to suppress behavior, but both factors must be simultaneously favorable to permit expression. This asymmetry mirrors the structure of biological social approach behavior, wherein both environmental safety and relational familiarity must be present for an organism to exhibit affiliative behavior, and the absence of either produces withdrawal regardless of the strength of the other. The minimum gate thus enforces a conservative behavioral policy that cannot be overridden by high values in either dimension alone, providing an inherent safety property that is architecturally guaranteed rather than learned or tuned.

[0047] In a further embodiment, the minimum gate may be replaced with an asymmetric gate for high-familiarity contexts. When the context coherence exceeds a high-familiarity threshold, the effective coherence may be computed as a weighted average of instantaneous and context coherence rather than a strict minimum, reflecting the resilience that deep familiarity provides against momentary perturbation. The weighting may be a function of the interaction count, such that more deeply familiar contexts provide greater dampening of momentary coherence drops.

5. Two-Dimensional Behavioral Phase Space

[0048] Referring now to FIG. 2, the effective coherence value and a concurrent tension value define a two-dimensional behavioral phase space. This phase space is partitioned into four behavioral quadrants, each producing a distinct behavioral profile.

FIG. 2 - Two-Dimensional Behavioral Phase Space

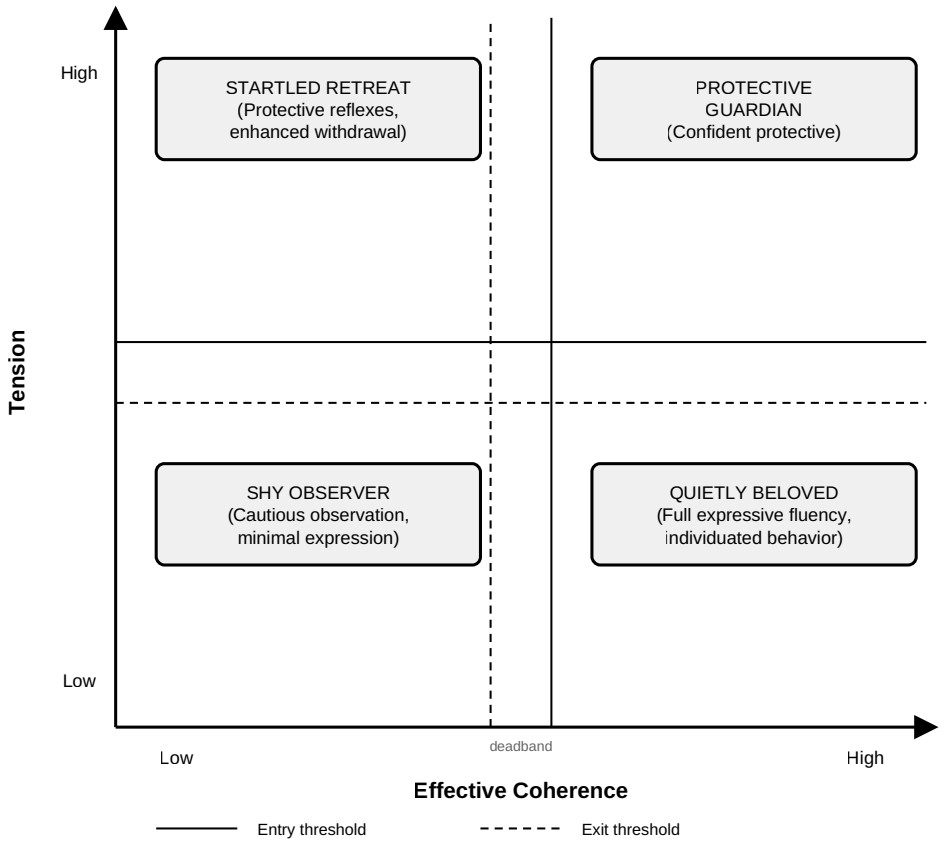

FIG. 2 — Two-dimensional behavioral phase space: four quadrants defined by effective coherence and tension axes.

[0049] The first quadrant, designated "Shy Observer," corresponds to low effective coherence and low tension. In this quadrant, the robot exhibits minimal behavioral expression, cautious observation, and reduced motor amplitude. The robot is in a learning state, accumulating information about an unfamiliar context.

[0050] The second quadrant, designated "Startled Retreat," corresponds to low effective coherence and high tension. In this quadrant, the robot exhibits protective reflexes combined with increased withdrawal distance. The robot retreats further and faster from perceived threats in unfamiliar contexts than it would in familiar ones.

[0051] The third quadrant, designated "Quietly Beloved," corresponds to high effective coherence and low tension. In this quadrant, the robot's full expressive range is unlocked. Personality parameters are expressed at full scale, producing individuated behavioral characteristics, idiosyncratic movement patterns, and rich communicative output. This quadrant represents the earned fluency state that is the developmental endpoint of the system.

[0052] The fourth quadrant, designated "Protective Guardian," corresponds to high effective coherence and high tension. In this quadrant, the robot acts with confidence derived from deep contextual familiarity while responding to an alarming stimulus. The robot may position itself protectively, emit warning signals calibrated to the specific household, or take defensive action informed by its accumulated knowledge of the context.

[0053] Quadrant boundaries are implemented with hysteresis using a Schmitt trigger pattern to prevent behavioral oscillation at boundary conditions. In a preferred embodiment, each quadrant has distinct entry and exit thresholds separated by a deadband. For example, entry into the Quietly Beloved quadrant may require coherence exceeding 0.65 and tension below 0.25, while exit requires coherence falling below 0.55 or tension exceeding 0.35. The deadband width (0.10 in this example) prevents flickering between behavioral modes when the robot's state is near a boundary.

6. Permeability Function

[0054] The behavioral output of the system is governed by a permeability function that maps the two-dimensional coherence-tension state to a behavioral envelope. The behavioral envelope defines the permissible range and intensity of the robot's outputs across all modalities, including but not limited to: motor movement amplitude and velocity, visual indicator (LED) color, intensity, and pattern, audio output volume, complexity, and tonal character, and verbal or narrative output verbosity, personality coloring, and emotional depth.

[0055] In a preferred embodiment, the permeability function operates as a continuous scaling function within each quadrant, with the quadrant determining the qualitative character of behavior and the exact position within the quadrant determining quantitative intensity.

[0055a] It is an architecturally significant and deliberate property of the invention that the system's entire output space is bounded to self-modulation of the robot's own behavioral expressiveness. The permeability function scales the amplitude, intensity, and complexity of the robot's existing behavioral repertoire. It does not generate novel actions, directives, or interventions targeted at humans or the external environment beyond the robot's own behavioral envelope. Specifically, the system contains no mechanism for: evaluative assessment or classification of humans or human behavior; assignment of categorical labels, threat levels, risk scores, or moral judgments to observed entities; generation of recommendations, decisions, or actions directed at humans based on the system's internal state; punitive, retaliatory, or corrective action in response to negative interaction events; or any output that modifies the external environment beyond the robot's own physical and communicative behavior.

[0055b] This constraint is not a policy layer, safety filter, or learned behavioral inhibition that could be overridden, fine-tuned, or removed by subsequent modification. It is a structural property of the system architecture: the permeability function accepts a two-dimensional input (effective coherence and tension) and produces a scalar behavioral envelope that modulates pre-existing output channels. There is no pathway in the system architecture through which internal state (including negative coherence, high tension, conflict signals, or low accumulated trust) can produce output directed at external entities rather than modulation of the robot's own expressiveness. The maximum behavioral consequence of any negative event or accumulation of negative events is complete behavioral withdrawal: zero motor amplitude, minimal visual indication, no verbal output. The system's failure mode under adversarial or traumatic conditions is silence and stillness, not aggression, retaliation, or evaluative judgment. This property holds for all possible values of all internal state variables, for all possible sensor inputs, and for all possible configurations of personality parameters, mixing matrices, suppression maps, and compiled routines, without exception.

[0055c] The four behavioral quadrants defined by the phase space (Shy Observer, Startled Retreat, Quietly Beloved, Protective Guardian) exhaustively enumerate the qualitative behavioral modes available to the system. No quadrant contains aggressive, punitive, evaluative, or externally-directed action. The Protective Guardian quadrant, which represents the highest-coherence, highest-tension state, produces confident protective positioning of the

robot's own body and heightened sensory alertness, not protective action directed at or against external agents. The system's relationship to humans is structurally asymmetric: humans modulate the robot's coherence through their interactions; the robot modulates only its own expressiveness in response. The human is the agent; the robot is the respondent. This asymmetry is not a design preference but a mathematical consequence of the architecture: coherence flows inward from interaction, expressiveness flows outward through the permeability function, and no pathway exists for evaluative judgment to flow outward toward the humans whose interactions produced the coherence.

6A. Manifold-Constrained Coherence Mixing

[0056] Referring now to FIG. 10, in a further embodiment, the system incorporates a manifold-constrained coherence mixing subsystem 135 that enables bounded cross-context coherence transfer while preserving the total coherence energy across all context streams. This subsystem addresses the limitation inherent in purely independent coherence accumulators: contexts that are closely related (e.g., the same room at different times of day, or the same person approaching from different directions) must either share a single accumulator (losing discriminatory capacity) or maintain entirely separate accumulators (losing the relational continuity that human social cognition naturally provides).

FIG. 10 - Manifold-Constrained Coherence Mixing Architecture

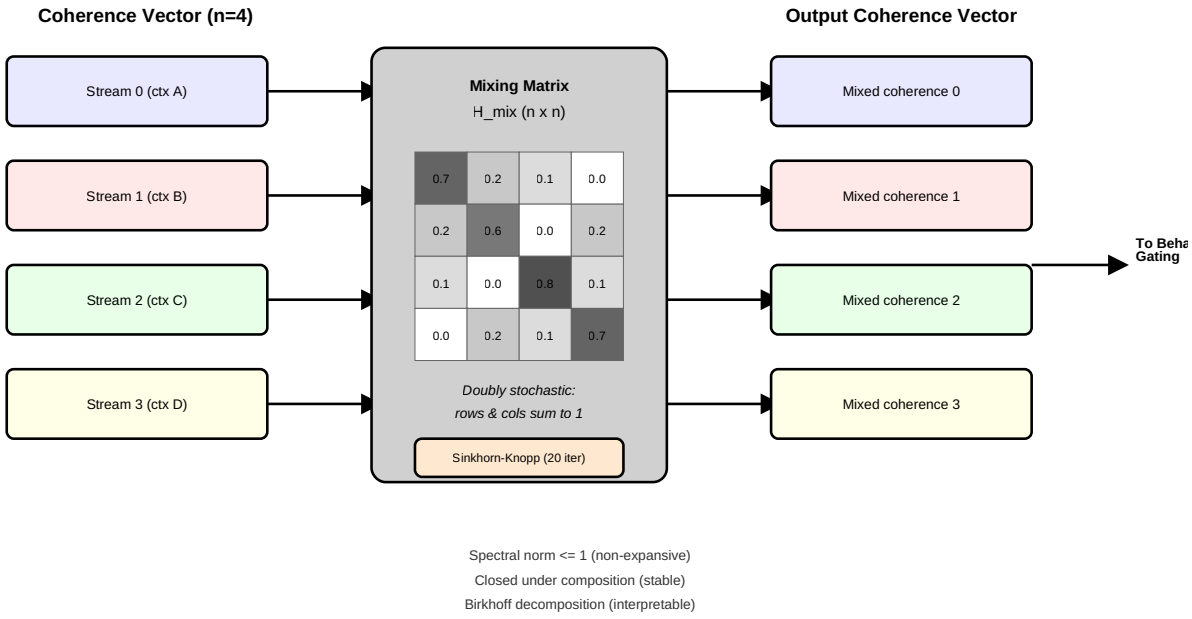

FIG. 10 — Manifold-constrained coherence mixing architecture and Birkhoff polytope projection.

[0057] The manifold-constrained coherence mixing subsystem organizes a set of n active coherence accumulators into a coherence vector c_l of dimension n, where n is the number of currently active context streams. At each processing tick, prior to the computation of effective coherence by the behavioral gating subsystem 130, the coherence vector is transformed by a mixing matrix H_mix of dimension n-by-n. The mixing matrix H_mix is constrained to be a doubly stochastic matrix: a matrix with non-negative entries where both rows and columns sum to 1.

[0058] The doubly stochastic constraint is enforced by projecting the unconstrained mixing parameters onto the Birkhoff polytope via the Sinkhorn-Knopp algorithm. Given an unconstrained parameter matrix M, the projection proceeds by: (a) applying an element-wise exponential to produce a positive matrix M_pos = exp(M); (b) alternately normalizing rows and columns of M_pos for a fixed number of iterations t_max (in a preferred embodiment, t_max = 20); (c) yielding a doubly stochastic matrix H_mix that is the closest projection of M_pos onto the Birkhoff polytope in the entropic sense.

[0059] The doubly stochastic constraint on H_mix confers three mathematical properties that are directly beneficial for social robot behavioral management:

[0060] First, norm preservation. The spectral norm of a doubly stochastic matrix is bounded by 1. This means the mixing operation is non-expansive: the coherence vector after mixing cannot have a larger norm than the coherence vector before mixing. In behavioral terms, the robot cannot amplify coherence through cross-context transfer. Trust cannot be manufactured; it can only be redistributed.

[0061] Second, compositional closure. The product of doubly stochastic matrices is itself doubly stochastic. This means that no matter how many mixing operations are composed over the operational life of the robot (potentially millions of processing ticks over months of operation), the composite mixing operation retains the non-expansive property. This provides a mathematical guarantee of long-term behavioral stability that engineering heuristics such as asymptotic ceilings and decay floors approximate but cannot formally guarantee.

[0062] Third, Birkhoff polytope geometry. By the Birkhoff-von Neumann theorem, every doubly stochastic matrix can be expressed as a convex combination of permutation matrices. In behavioral terms, the robot's cross-context coherence mixing is always interpretable as a weighted blend of discrete context-swapping operations. This provides geometric interpretability: the mixing behavior at any moment can be decomposed into "how much of context A's coherence is being treated as if it were context B's coherence," with weights that are non-negative and sum appropriately.

[0063] The mixing matrix H_mix comprises two components, following the parameterization approach of manifold-constrained hyper-connections in deep learning:

[0064] A static component b_mix, which is a learned bias matrix representing the baseline cross-context relationships that do not depend on current sensor input. This component captures stable relationships such as "morning kitchen context and midday kitchen context should share some coherence."

[0065] A dynamic component computed as alpha_mix multiplied by a learned projection of the current sensor state, where alpha_mix is a gating factor initialized to a small value (in a preferred embodiment, 0.01). The dynamic component enables the mixing to be modulated by the current environmental context: the degree of cross-context transfer may vary depending on what the robot is currently sensing.

[0066] The initialization of the gating factor alpha_mix near zero produces a critical behavioral property: at the beginning of the robot's operational life, the mixing matrix is dominated by the static bias, which is initialized near the identity matrix. This means the robot begins with virtually no cross-context coherence transfer. Each context stream maintains its coherence independently, producing the maximally reserved, shy behavior that is the developmental starting point of the CCF system.

[0067] As the robot accumulates interaction experience and the deliberative processing unit 220 optimizes the mixing matrix parameters (described in Part II), the gating factor alpha_mix and the bias b_mix evolve, permitting increasingly sophisticated cross-context transfer. The transition from shy to relationally nuanced behavior is not programmed; it is architecturally earned through a combination of parameter learning and manifold-constrained evolution.

[0068] The near-zero entries of the converged mixing matrix H_mix correspond directly to context boundaries: pairs of contexts between which the robot has learned that coherence transfer is inappropriate. Conversely, large entries indicate contexts between which the robot has learned that coherence may be safely shared. This provides a continuous, real-time representation of context boundary topology that complements (and in some embodiments replaces) the discrete graph min-cut computation described in Section 7.

[0069] Referring now to FIG. 11, the evolution of the mixing matrix over the operational life of the robot follows a characteristic developmental trajectory: (a) initial state: near-identity matrix, minimal cross-context transfer, predominantly reserved behavior across all contexts; (b) early operation: small off-diagonal entries emerge as the robot begins to learn which contexts are related, producing cautious partial transfer; (c) mature operation: the matrix

exhibits clear block-diagonal structure corresponding to coherence domains (groups of contexts that share coherence) separated by near-zero entries corresponding to context boundaries, with the block structure matching the partition that would be obtained by applying min-cut to the relational graph.

FIG. 11 - Mixing Matrix Developmental Evolution

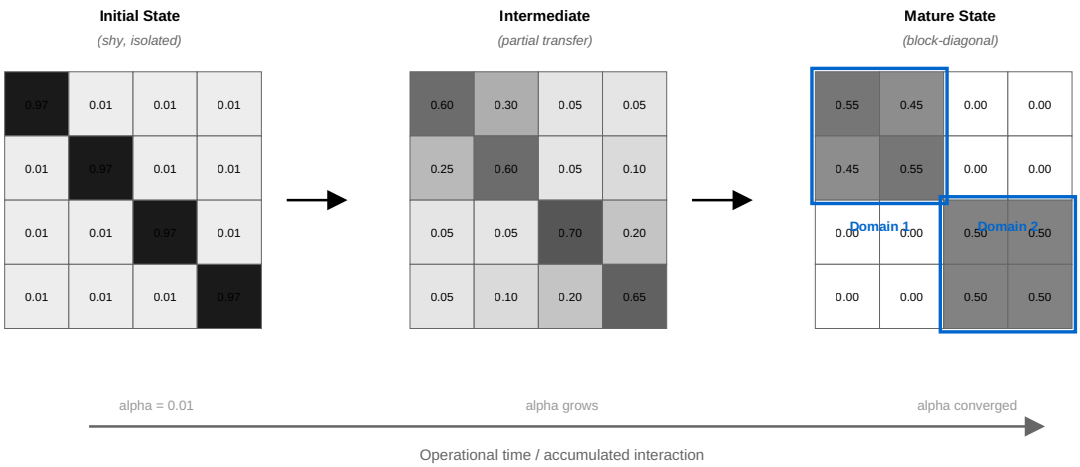

FIG. 11 — Evolution of the mixing matrix over the operational life of the robot.

[0070] In the manifold-constrained embodiment, the effective coherence for behavioral gating may be computed from the full coherence vector after mixing, rather than from a single scalar accumulator. This enables a multi-dimensional behavioral phase space wherein the robot's behavior is modulated by its simultaneous relationship to multiple context domains, producing richer behavioral differentiation than the two-dimensional coherence-tension phase space of the scalar embodiment.

7. Automatic Context Boundary Discovery via Graph Min-Cut

[0071] Referring now to FIG. 4, the context boundary discovery subsystem 140 addresses the context granularity problem: the question of which combinations of sensor features should be treated as the same context (sharing a coherence accumulator) and which should be treated as distinct contexts (maintaining separate accumulators).

FIG. 4 - Context Boundary Discovery: Graph Min-Cut and Mixing Matrix

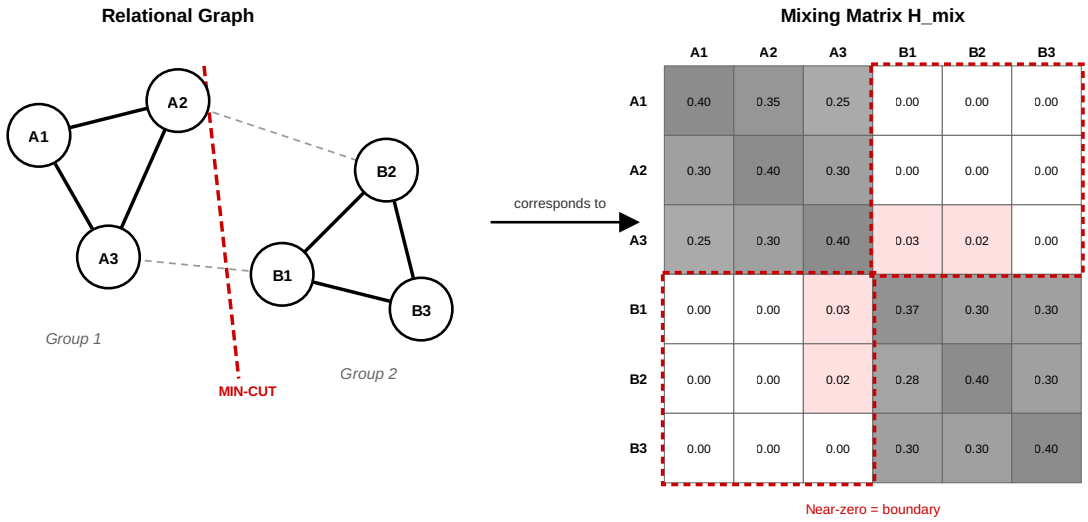

FIG. 4 — Relational graph construction and graph min-cut for context boundary discovery.

[0072] The subsystem constructs a relational graph from accumulated interaction episodes. Each interaction episode is a record comprising a context key, an outcome classification (positive or negative), and a trajectory vector representing the time-series pattern of tension, coherence, and energy values over the duration of the interaction.

[0073] In the relational graph, nodes represent observed context keys and edges represent behavioral similarity between contexts. Edge weights are computed as the cosine similarity of trajectory vectors between contexts, thresholded at a minimum similarity value below which contexts are considered unrelated. This produces a graph where clusters of tightly connected nodes represent contexts that produce similar behavioral patterns and may share coherence, while weakly connected or disconnected clusters represent genuinely different social situations requiring independent coherence tracking.

[0074] A graph min-cut algorithm, in a preferred embodiment the Stoer-Wagner algorithm, is applied to the relational graph to find the minimum-weight set of edges whose removal partitions the graph into disconnected components. Each resulting partition defines a coherence group: contexts within the same partition share a single coherence accumulator, while contexts in different partitions maintain independent accumulators.

[0075] The min-cut computation may be applied recursively (N-way bisection) to discover multiple coherence groups. The recursion terminates when the cut weight exceeds a personality-dependent threshold, meaning that further partitioning would sever too much behavioral similarity. A robot with high curiosity drive tolerates more context separation (producing more coherence groups and finer discrimination), while a cautious robot prefers fewer groups (more coherence sharing, slower to discriminate).

[0076] Cross-partition connections identified by the min-cut algorithm, herein termed "bleeding edges," represent behavioral leakage between coherence domains. The magnitude of these bleeding edges provides a quantitative signal indicating how strongly the system recommends merging or separating specific contexts.

[0077] The relational graph is rebuilt and the min-cut is recomputed periodically, at intervals scaling with the total number of observed contexts. This means context boundaries are not fixed at design time but emerge from accumulated experience and evolve as the robot encounters new situations.

[0078] In the manifold-constrained embodiment, the graph min-cut computation and the mixing matrix structure provide complementary representations of context boundary topology. The min-cut provides a discrete partition suitable for the deliberative unit's periodic

consolidation, while the mixing matrix provides a continuous, real-time representation suitable for the reflexive unit's per-tick processing. During consolidation, the deliberative unit may cross-validate the min-cut partition against the mixing matrix structure, using agreement between the two representations as a confidence signal for context boundary quality.

7A. Unexpected Emergent Properties

[0078a] The CCF system, when operated over extended periods, produces several emergent behavioral phenomena that are not explicitly programmed into any individual component and would not be predicted by a person of ordinary skill in the art from examination of the components in isolation. These emergent properties constitute unexpected results of the specific combination of context-keyed accumulation, minimum-gate behavioral gating, manifold-constrained mixing, graph min-cut boundary discovery, and dual-process cognitive architecture.

[0078b] First, context-specific shyness without shyness programming. The system produces behavior that human observers consistently describe as "shy" or "cautious" in unfamiliar contexts, without any component being designed or parameterized to produce shyness. The shy behavior emerges from the interaction of the minimum gate (which suppresses expressiveness when accumulated coherence is low) with the zero-initialization of coherence accumulators for new contexts. No personality parameter controls shyness directly. A robot with high curiosity drive and high recovery speed still exhibits initial shyness in novel contexts because the minimum gate enforces the architectural invariant regardless of personality parameterization. This is unexpected because existing approaches achieve reserved behavior through explicit shyness or introversion parameters, and a skilled practitioner would expect that removing such parameters would produce uniformly expressive behavior.

[0078c] Second, spontaneous convergence of mixing matrix structure and graph min-cut partition. In the manifold-constrained embodiment, the doubly stochastic mixing matrix, which is optimized through gradient-based parameter updates during consolidation, independently converges to a block-diagonal structure whose block boundaries correspond to the partition computed by the Stoer-Wagner min-cut algorithm on the relational graph. These two computations operate on different representations (continuous matrix parameters versus discrete graph edges), use different algorithms (Sinkhorn-Knopp projection versus combinatorial graph partitioning), and are optimized for different objectives (coherence prediction accuracy versus minimum-weight edge removal). Their convergence to the same boundary topology is an emergent property of the underlying structure of social context relationships and would not be predicted from examination of either algorithm in isolation.

[0078d] Third, observable hesitation as perceived cognitive depth. When the reflexive and deliberative processing units produce divergent behavioral assessments (described in detail in Section 12), the system enters a conflict resolution state that produces a visible pause in behavioral output. Human observers do not interpret this pause as a system delay or error. Instead, they consistently interpret it as evidence that the robot is "thinking" or "considering" its response, attributing cognitive depth and internal experience to the robot. This attribution of inner life from a mechanism that is fundamentally a classification conflict between two processing pathways is an emergent social phenomenon that would not be predicted by a skilled practitioner designing a dual-process architecture for computational efficiency.

[0078e] Fourth, asymmetric trust resilience proportional to interaction depth. The interaction-count-proportional decay floor (Section 3) combined with the asymptotic growth function produces a trust resilience profile wherein a robot that has interacted positively with a context over hundreds of episodes recovers from a negative event (returning to near-previous coherence levels) orders of magnitude faster than a robot that has interacted with the same context over only a few episodes. This produces the emergent behavioral property of "deep trust": long-familiar contexts exhibit a qualitative robustness to perturbation that short-familiar contexts do not, even when both contexts have the same current coherence value. This qualitative distinction between shallow and deep familiarity at identical coherence levels is an emergent property of the interaction between the floor function and the growth function and is not programmed as a separate mechanism.

[0078f] Fifth, developmental trajectory without developmental programming. Over extended operation (weeks to months), the system traverses a characteristic developmental arc from reserved and undifferentiated behavior across all contexts, through a period of increasing context discrimination and selective expressiveness, to a mature state with rich context-specific behavioral repertoires, compiled habitual routines, and sophisticated cross-context transfer patterns (in the manifold-constrained embodiment). This developmental trajectory resembles biological social development in its broad structure, but is not implemented as a developmental program. No component of the system tracks developmental stage, implements developmental transitions, or encodes a target mature state. The trajectory emerges from the accumulation of coherence, the compilation of routines, the evolution of the mixing matrix, and the refinement of context boundaries through ongoing interaction. A skilled practitioner designing the individual components would not predict this specific developmental trajectory from the component specifications.

PART II: DUAL-PROCESS COGNITIVE ARCHITECTURE

8. Architectural Overview and Biological Analogue

[0079] Referring now to FIG. 6, the dual-process cognitive architecture 200 comprises a reflexive processing unit 210 and a deliberative processing unit 220 connected by bidirectional modulation pathways. This architecture is analogous to the dual-process structure observed in biological neural systems, wherein subcortical structures (amygdala, brainstem, cerebellum) execute fast reflexive processing while cortical structures (prefrontal cortex, hippocampus) execute slower deliberative processing, with each system capable of modulating, inhibiting, or overriding the other.

FIG. 6 - Bidirectional Modulation Pathways

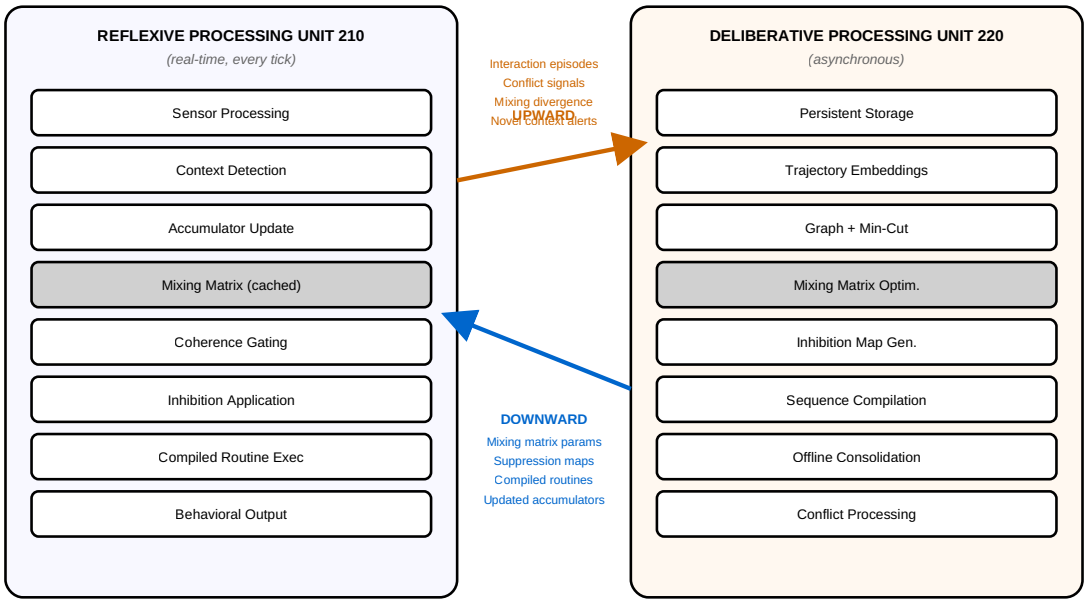

FIG. 6 — Bidirectional modulation pathways of the dual-process cognitive architecture.

[0080] The reflexive processing unit 210 executes at every processing tick and is responsible for: sensor reading and quantization, context key construction, coherence accumulator lookup and update, application of the cached mixing matrix (in the manifold-constrained embodiment), effective coherence computation via the minimum gate (or asymmetric gate), behavioral phase space quadrant determination, application of learned inhibition maps received from the deliberative unit, execution of compiled behavioral sequences received from the deliberative unit, and generation of behavioral output through the permeability function.

[0081] The deliberative processing unit 220 executes asynchronously and is responsible for: persistent storage and retrieval of coherence accumulators and interaction episodes, trajectory embedding computation, relational graph construction and min-cut computation for context boundary discovery, mixing matrix parameter optimization (in the manifold-constrained embodiment), generation and updating of learned inhibition maps, compilation of frequently-repeated behavioral sequences into reflexive routines, offline consolidation of accumulated experience during idle periods, and processing of conflict signals received from the reflexive unit.

[0082] The bidirectional modulation between units is characterized by two pathways. The downward pathway (deliberative to reflexive) carries: updated coherence accumulators with revised context boundaries, updated mixing matrix parameters (in the manifold-constrained embodiment), learned inhibition maps specifying context-specific stimulus suppression, compiled behavioral sequences for autonomous reflexive execution, and updated prior expectations for anticipatory processing. The upward pathway (reflexive to deliberative) carries: raw interaction episodes for persistent storage and graph construction, novel context notifications when previously unobserved context keys are detected, mixing matrix divergence signals when the cached mixing matrix produces unexpected coherence transfer patterns, and classification conflict signals when the reflexive and deliberative units produce divergent behavioral assessments.

9. Context-Specific Learned Inhibition (Startle Override)

[0083] The deliberative processing unit 220 generates context-specific stimulus suppression maps and pre-loads them to the reflexive processing unit 210, enabling the reflexive unit to dampen reflexive responses to stimuli that have been learned as benign in the current context.

[0084] A stimulus suppression map is a data structure associating context keys with stimulus-response modification rules. Each rule specifies: a stimulus pattern (e.g., a loudness spike exceeding a threshold), a context key or set of context keys in which the suppression applies, a

suppression factor (a scalar between 0.0 and 1.0 that attenuates the reflexive response), and an experience count (the number of times this stimulus has been observed without negative consequence in this context).

[0085] The deliberative processing unit constructs stimulus suppression maps by analyzing accumulated interaction episodes. When a stimulus pattern that initially triggered a strong reflexive response (e.g., a loud noise causing startle and coherence drop) has been observed a configurable number of times in a specific context without subsequent negative outcome (no collision, no sustained tension increase), the deliberative unit generates a suppression rule for that stimulus-context pair and transmits it to the reflexive unit.

[0086] The reflexive processing unit applies suppression maps prior to the coherence update step. When a stimulus is detected that matches a suppression rule for the current context, the reflexive response (tension increase, coherence decrease) is attenuated by the suppression factor. This produces the observable behavior of a robot that flinches at an unfamiliar loud noise in a new context but remains calm when the same noise occurs in a familiar context where it has been learned as benign. The suppression is context-specific: the same stimulus in a different context, where it has not been learned as benign, produces the full unsuppressed reflexive response.

[0087] This mechanism is analogous to the biological process of cortical inhibition of subcortical startle responses, wherein the prefrontal cortex modulates amygdala-mediated fear responses based on contextual learning and threat assessment.

[0088] In the manifold-constrained embodiment, suppression maps may be partially transferred between contexts connected by high-weight entries in the doubly stochastic mixing matrix. When context A has a well-established suppression rule for a stimulus and context B shares substantial coherence with context A (as indicated by a high mixing matrix entry), the suppression rule may be applied with reduced suppression factor in context B, producing an attenuation of startle response that is proportional to the learned cross-context relationship. This provides a principled mechanism for partial generalization of learned inhibition without the all-or-nothing transfer that would occur if suppression maps were shared globally.

10. Behavioral Sequence Compilation (Habit Formation)

[0089] Referring now to FIG. 7, the deliberative processing unit 220 implements a behavioral sequence compilation mechanism that transfers frequently-repeated deliberative behavioral sequences to the reflexive processing unit 210 for autonomous execution.

FIG. 7 - Behavioral Sequence Compilation Lifecycle

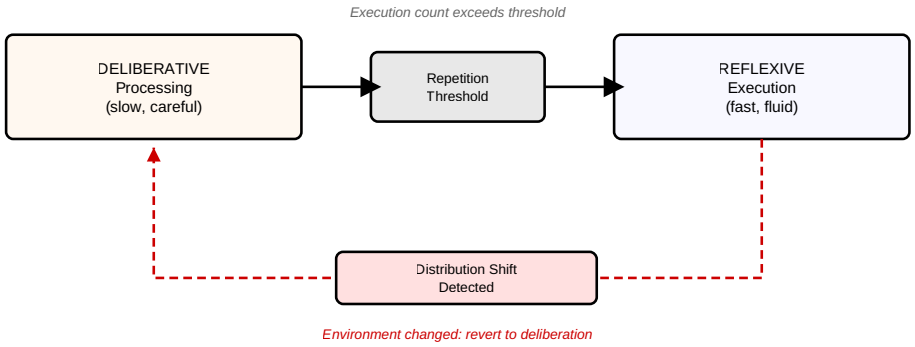

FIG. 7 — Habit compilation lifecycle: deliberative processing to compiled reflexive execution.

[0090] A behavioral sequence is a temporally ordered series of motor, visual, and audio output commands that the deliberative unit has computed in response to a specific context and sensor pattern. The deliberative unit maintains a repertoire of behavioral sequences, each associated with: a trigger condition (context key and sensor pattern), a sequence of output commands, an execution count (the number of times this sequence has been successfully executed), and a distribution signature (a statistical summary of the sensor conditions under which successful executions have occurred).

[0091] When the execution count for a behavioral sequence exceeds a compilation threshold (which may be personality-modulated, with higher curiosity drive requiring fewer repetitions for compilation), the deliberative unit compresses the sequence into a compiled routine and transmits it to the reflexive unit. The compiled routine includes the trigger condition, the compressed output sequence, and the distribution signature.

[0092] The reflexive processing unit maintains a cache of compiled routines. At each processing tick, when the current context and sensor pattern match a compiled routine's trigger condition, the reflexive unit executes the compiled routine autonomously without consulting the deliberative unit. This produces faster, more fluid behavioral execution in familiar situations.

[0093] Critically, the reflexive unit monitors the current sensor conditions against the compiled routine's distribution signature during execution. When the current conditions deviate from the distribution signature beyond a configurable threshold (indicating that the environment has changed in a way not captured by the routine's training history), the reflexive unit halts autonomous execution and generates a distribution shift signal on the upward pathway to the deliberative unit, which resumes deliberative control of the behavior. This mechanism is analogous to the biological process of skill acquisition and breakdown: practiced behaviors become automatic ("muscle memory") but revert to conscious deliberation when environmental conditions change.

[0094] The compilation lifecycle thus follows a predictable arc: novel behavior begins in the deliberative unit (slow, careful, context-checking), is repeated and refined through interaction, crosses the compilation threshold, and is transferred to the reflexive unit (fast, fluid, autonomous). If conditions change, it returns to the deliberative unit for re-evaluation and potential recompilation. This cycle may repeat indefinitely as the robot's environment evolves.

11. Offline Consolidation Cycles

[0095] Referring now to FIG. 8, the deliberative processing unit 220 implements offline consolidation cycles during periods when the robot is idle, charging, or otherwise not engaged in active interaction. During consolidation, the deliberative unit performs the following operations.

FIG. 8 - Offline Consolidation Cycle

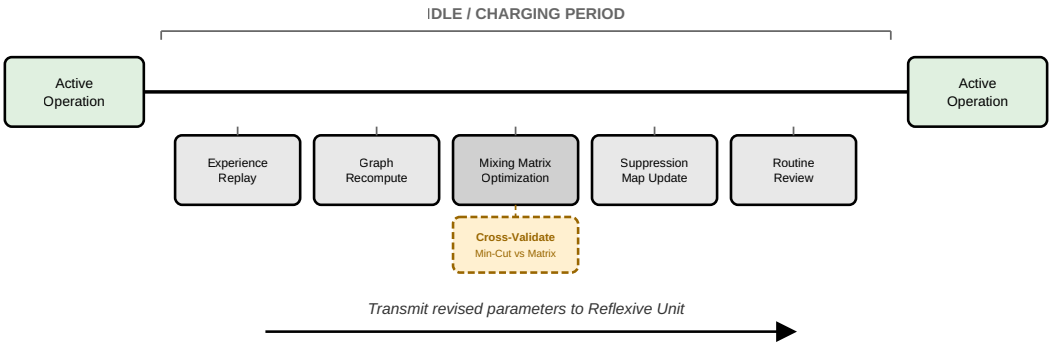

FIG. 8 — Consolidation cycle timeline during idle period.

[0096] Experience replay: Accumulated interaction episodes are replayed and re-analyzed. Trajectory embeddings may be recomputed with updated parameters. Patterns that were not detectable in real-time (due to processing constraints or insufficient data) may become apparent when episodes are analyzed in batch.

[0097] Relational graph recomputation: The relational graph is rebuilt from the full set of accumulated episodes and the min-cut algorithm is reapplied. This may produce revised coherence group boundaries, merging contexts that were previously separate (if accumulated evidence shows they are behaviorally equivalent) or splitting contexts that were previously merged (if accumulated evidence reveals behavioral divergence).

[0098] Mixing matrix optimization: In the manifold-constrained embodiment, the deliberative unit optimizes the parameters of the mixing matrix (both the static bias component and the dynamic projection parameters) based on accumulated interaction data. The optimization objective is to minimize a loss function comprising: (a) a cross-context prediction error term that rewards mixing matrices producing accurate predictions of coherence dynamics across related contexts; (b) a sparsity term that encourages near-zero entries for unrelated context pairs; and (c) a stability term that penalizes mixing matrices producing large changes to the coherence vector norm. The optimized parameters are projected onto the Birkhoff polytope via Sinkhorn-Knopp and transmitted to the reflexive unit. The mixing matrix structure after optimization is cross-validated against the min-cut partition as a boundary quality check.

[0099] Stimulus suppression map updating: Suppression rules are reviewed against accumulated evidence. Rules may be strengthened (higher suppression factor) if the benign status of a stimulus has been further confirmed, weakened if subsequent evidence suggests the stimulus is not reliably benign, or created for newly identified stimulus-context pairs.

[0100] Compiled routine review: Compiled behavioral sequences are reviewed against accumulated outcomes. Routines with degrading success rates may be decompiled and returned to deliberative control for relearning.

[0101] Following a consolidation cycle, the deliberative unit transmits updated coherence accumulators, revised context boundaries, optimized mixing matrix parameters, updated suppression maps, and revised compiled routines to the reflexive unit via the downward pathway. The robot thus resumes active operation with a refined relational understanding that reflects not only new experience but also the reorganization and reinterpretation of existing experience.

[0102] This mechanism is functionally analogous to memory consolidation during sleep in biological neural systems, wherein hippocampal memory traces are replayed and reorganized, strengthening important associations and pruning irrelevant ones. The robot "wakes up" from a consolidation cycle with potentially different relational categories, mixing patterns, and behavioral calibrations than it had before the cycle, despite no new external interaction having occurred.

12. Classification Conflict Resolution

[0103] Referring now to FIG. 9, a distinctive feature of the dual-process architecture is the handling of classification conflicts: situations where the reflexive processing unit 210 and the deliberative processing unit 220 produce divergent behavioral assessments. This phenomenon, herein designated the "Conflicted Classification Problem," arises when the reflexive unit's real-time sensor assessment indicates one behavioral quadrant while the deliberative unit's contextual knowledge and accumulated experience indicates a different quadrant. As an illustrative analogy, this is comparable to a classification system that receives input matching multiple incompatible categories simultaneously, producing a sustained state of irresolution rather than a clean categorical output.

FIG. 9 - Classification Conflict Resolution

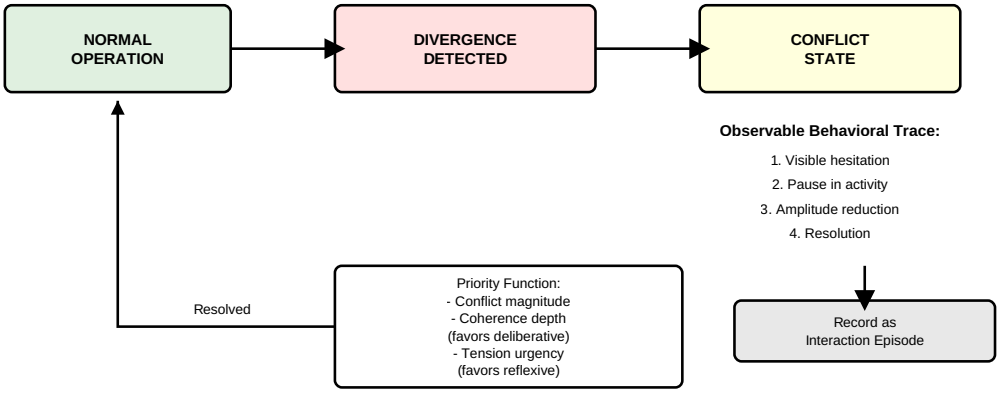

FIG. 9 — Classification conflict resolution: hesitation trace and resolution pathway.

[0104] In existing robotic systems, such conflicts are typically resolved invisibly through a fixed priority hierarchy (e.g., reflexive always wins, or deliberative always wins) or through simple averaging. The present invention instead treats the conflict itself as a meaningful behavioral signal.

[0105] When the reflexive unit and deliberative unit produce divergent behavioral assessments, the system enters a conflict resolution state characterized by an observable behavioral trace: a visible hesitation, a pause in ongoing activity, a brief reduction in behavioral amplitude across all output modalities, followed by a resolution in which one assessment prevails. The prevailing assessment is determined by a context-sensitive priority function that considers the magnitude of the conflict (how far apart the two assessments are), the coherence depth of the current context (deeper familiarity gives the deliberative unit more weight), and the tension urgency of the current moment (higher tension gives the reflexive unit more weight).

[0106] In the manifold-constrained embodiment, the magnitude of conflict may be quantified as the distance between the reflexive unit's mixing matrix state (as cached) and the deliberative unit's current mixing matrix state (as updated from accumulated experience), measured in the metric space of doubly stochastic matrices. A large distance indicates that the reflexive unit is operating with an outdated model of cross-context relationships, and the hesitation behavior is correspondingly more pronounced.

[0107] The observable behavioral trace of conflict resolution is not a deficiency or a latency artifact. It is a deliberate design feature that produces the appearance of internal deliberation: the robot visibly encounters a situation that gives it pause, considers, and then acts. Human observers interpret this hesitation-and-resolution pattern as evidence of internal depth, cognitive complexity, and authentic engagement with the environment. This contributes significantly to the perception of the robot as a being with genuine personality rather than a deterministic automaton executing pre-programmed responses.

[0108] Following resolution, the conflict event is recorded as an interaction episode on the upward pathway, including both the reflexive and deliberative assessments and the resolution outcome. This record informs future deliberative processing: repeated conflicts between the same reflexive and deliberative assessments in the same context may lead to the generation of a stimulus suppression map (if the deliberative assessment consistently prevails) or the recompilation of a behavioral sequence (if the conflict reveals that an existing compiled routine is no longer appropriate for the evolving context).

13. Degraded Mode and Reconnection

[0109] In a degraded mode, when the deliberative processing unit 220 is unavailable due to connectivity loss, resource exhaustion, or maintenance, the reflexive processing unit 210 continues to operate autonomously using its locally cached coherence accumulators, suppression maps, compiled routines, and mixing matrix parameters (in the manifold-constrained embodiment). In this mode, the effective coherence is computed from instantaneous sensor-derived coherence only (equivalent to prior art global coherence), as context-specific accumulators cannot be updated without the deliberative unit's persistent storage. The cached mixing matrix continues to apply cross-context transfer based on previously learned relationships, providing partial relational continuity even during degraded operation.

[0110] Upon reconnection of the deliberative processing unit, the reflexive unit transmits all interaction episodes accumulated during the degraded period. The deliberative unit integrates these episodes into its persistent store, recomputes context boundaries and suppression maps as needed, re-optimizes the mixing matrix parameters, and resumes full bidirectional modulation. The system thus degrades gracefully to a simpler but functional behavioral model and recovers seamlessly when the deliberative capacity is restored.

REFERENCES

- Moshkina, L. and Arkin, R. "Human perspective on affective robotic behavior: A longitudinal study." Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 1444-1451, 2005.
- Lee, K. M., Peng, W., Jin, S. A., and Yan, C. "Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction." Journal of Communication, 56(4):754-772, 2006.
- Tapus, A. and Mataric, M. "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance." AAAI Spring Symposium on Emotion, Personality, and Social Behavior, 2008.
- Joosse, M., Lohse, M., Perez, J. G., and Evers, V. "What you do is who you are: The role of task context in perceived social robot personality." Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pages 2134-2139, 2013.
- Xie, Z., Wei, Y., Cao, H., et al. "mHC: Manifold-Constrained Hyper-Connections." DeepSeek-AI, arXiv:2512.24880v2, January 2026.

- Zhu, D., Huang, H., Huang, Z., et al. "Hyper-Connections." arXiv:2409.19606, 2024.
- Sinkhorn, R. and Knopp, P. "Concerning nonnegative matrices and doubly stochastic matrices." Pacific Journal of Mathematics, 21(2):343-348, 1967.
- Stoer, M. and Wagner, F. "A simple min-cut algorithm." Journal of the ACM, 44(4): 585-591, 1997.
- He, K., Zhang, X., Ren, S., and Sun, J. "Identity mappings in deep residual networks." European Conference on Computer Vision, pages 630-645, 2016.

CLAIMS

What is claimed is:

1. A system for managing behavioral expressiveness of an autonomous robot, the system comprising: a context detection subsystem configured to construct a composite context key from a plurality of sensor signals, the composite context key representing a situation fingerprint characterizing the robot's current environmental and social context; a coherence accumulation subsystem comprising a plurality of coherence accumulators, each coherence accumulator associated with a distinct context key and independently tracking an accumulated coherence value representing relational familiarity for the associated context, wherein each coherence accumulator is updated independently based on interaction outcomes occurring within the associated context; a behavioral gating subsystem configured to compute an effective coherence value as a function of an instantaneous coherence derived from current sensor stability and the accumulated coherence value for the current context key, wherein the effective coherence value governs the behavioral expressiveness of the robot; and a behavioral output interface configured to modulate the robot's behavioral outputs based on the effective coherence value.

2. The system of claim 1, wherein the effective coherence value is computed as the minimum of the instantaneous coherence and the accumulated coherence value for the current context key, thereby enforcing an architectural invariant that the robot cannot express more behavioral richness than the lesser of its current environmental stability and its accumulated relational familiarity with the present context.

3. The system of claim 1, wherein each coherence accumulator further comprises an interaction count tracking the number of interaction episodes recorded in the associated

context, and wherein the coherence value decays toward a floor value that is proportional to the interaction count, such that contexts with deeper interaction histories maintain higher minimum coherence values and are more resilient to coherence loss from isolated negative events.

4. The system of claim 1, wherein positive accumulation of the coherence value employs an asymptotic growth function computed as a base increment rate multiplied by a factor of (1.0 minus the current coherence value), producing rapid initial gains and diminishing increments as the coherence value approaches a maximum.

5. The system of claim 1, wherein the coherence accumulation dynamics are modulated by a set of personality parameters including at least a recovery speed parameter governing the rate of positive coherence accumulation and a startle sensitivity parameter governing the rate of negative coherence accumulation, wherein the personality parameters modulate the rate at which coherence is earned or lost but do not modify the structural requirement that coherence must be earned through accumulated interaction.

6. The system of claim 1, further comprising a two-dimensional behavioral phase space defined by the effective coherence value and a concurrent tension value, the phase space partitioned into a plurality of behavioral quadrants each producing a distinct behavioral profile, wherein transitions between quadrants employ hysteresis with separate entry and exit thresholds to prevent behavioral oscillation at boundary conditions.

7. The system of claim 6, wherein the plurality of behavioral quadrants comprises: a first quadrant corresponding to low effective coherence and low tension, producing cautious observation behavior; a second quadrant corresponding to low effective coherence and high tension, producing protective retreat behavior with enhanced withdrawal; a third quadrant corresponding to high effective coherence and low tension, producing full expressive fluency with individuated behavioral characteristics; and a fourth quadrant corresponding to high effective coherence and high tension, producing confident protective behavior informed by accumulated contextual familiarity.

8. The system of claim 1, wherein the composite context key is formed by quantizing each sensor signal into discrete feature levels and concatenating the quantized feature levels, the sensor signals comprising at least two of: ambient light level, ambient sound level, proximity sensor pattern, accelerometer-derived motion classification, gyroscope-derived orientation classification, and temporal period classification.

9. A method for automatic discovery of context boundaries in a social robot system employing context-keyed coherence accumulators, the method comprising: accumulating a plurality of interaction episodes, each interaction episode comprising a context key identifying an environmental and social context, and a trajectory vector representing a time-series pattern of internal state variables over the duration of the interaction; constructing a relational graph wherein nodes represent observed context keys and edges are weighted by behavioral similarity between contexts, the behavioral similarity computed as a similarity metric between trajectory vectors of contexts; applying a graph min-cut algorithm to the relational graph to partition the graph into a plurality of coherence groups, wherein contexts within the same coherence group share a coherence accumulator and contexts in different coherence groups maintain independent accumulators; and periodically recomputing the graph partition as new interaction episodes are accumulated, whereby context boundaries emerge from accumulated experience and evolve over the operational life of the robot.

10. The method of claim 9, wherein the graph min-cut algorithm comprises the Stoer-Wagner algorithm, and wherein the min-cut is applied recursively as N-way bisection, terminating when a cut weight exceeds a threshold, the threshold being modulated by a personality parameter such that higher curiosity produces finer context discrimination.

11. The method of claim 9, further comprising identifying cross-partition connections in the min-cut result as bleeding edges representing behavioral leakage between coherence domains, wherein the magnitude of bleeding edges provides a quantitative signal for context boundary quality assessment.

12. The method of claim 9, wherein the trajectory vectors are computed from time-series values of tension, coherence, and energy internal state variables, and wherein the similarity metric is cosine similarity thresholded at a minimum value below which contexts are considered unrelated.

13. The method of claim 9, wherein early in the operational life of the robot, the relational graph yields few coherence groups producing predominantly reserved behavior, and after extended operation, the graph naturally partitions into a plurality of distinct relational domains with earned expressiveness in familiar domains and continued reserved behavior in novel domains.

14. A method for gating behavioral expressiveness of an autonomous robot, the method comprising: maintaining a plurality of context-keyed coherence accumulators, each accumulator independently tracking accumulated relational familiarity for a distinct environmental and social context detected by the robot; at each processing cycle, detecting a

current context from sensor inputs and retrieving the coherence accumulator associated with the detected context; computing an instantaneous coherence from current sensor stability; computing an effective coherence as a function that prevents the effective coherence from exceeding the accumulated context coherence, regardless of instantaneous environmental conditions; and modulating the robot's behavioral outputs through a permeability function that maps the effective coherence and a concurrent tension value to a behavioral envelope defining permissible output range and intensity across a plurality of output modalities.

15. The method of claim 14, wherein the function that prevents the effective coherence from exceeding the accumulated context coherence is a minimum function, such that the effective coherence equals the lesser of the instantaneous coherence and the accumulated context coherence.

16. The method of claim 14, wherein when the accumulated context coherence exceeds a high-familiarity threshold, the effective coherence is computed as a weighted average of instantaneous coherence and accumulated context coherence rather than a strict minimum, the weighting being a function of interaction count, thereby providing resilience against momentary perturbation in deeply familiar contexts.

17. The method of claim 14, wherein the plurality of output modalities comprises at least two of: motor movement amplitude and velocity, visual indicator color and pattern, audio output volume and tonal character, and verbal output verbosity and personality expression intensity.

18. A system for manifold-constrained cross-context coherence transfer in an autonomous social robot, the system comprising: a plurality of context-keyed coherence accumulators organized into a coherence vector of dimension n, each element of the coherence vector tracking accumulated relational familiarity for a distinct environmental and social context; a mixing matrix of dimension n-by-n constrained to be a doubly stochastic matrix via projection onto the Birkhoff polytope, the mixing matrix governing bounded cross-context coherence transfer such that coherence accumulated in one context may be partially transferred to related contexts subject to the constraint that total coherence energy across all context streams is conserved; and a behavioral gating subsystem that computes effective coherence from the coherence vector after application of the mixing matrix, and modulates the robot's behavioral outputs based on the effective coherence.

19. The system of claim 18, wherein the doubly stochastic constraint is enforced by projecting unconstrained mixing parameters onto the Birkhoff polytope via the Sinkhorn-Knopp algorithm comprising iterative alternating row and column normalization of a positive matrix until convergence to a doubly stochastic matrix.

20. The system of claim 18, wherein the mixing matrix comprises a static component representing baseline cross-context relationships independent of current sensor input, and a dynamic component computed from a learned projection of current sensor state modulated by a gating factor, the gating factor being initialized to a value near zero such that the robot begins operation with minimal cross-context coherence transfer and earns increased transfer capacity through accumulated interaction experience.

21. The system of claim 18, wherein the spectral norm of the mixing matrix is bounded by 1, ensuring that cross-context coherence transfer is non-expansive and the robot cannot amplify coherence through transfer, and wherein the set of doubly stochastic matrices is closed under composition, ensuring that the composite mixing operation across any number of processing cycles retains the non-expansive property and provides a mathematical guarantee of long-term behavioral stability.

22. The system of claim 18, wherein near-zero entries of the mixing matrix identify context boundaries between coherence domains, and wherein the context boundary topology represented by the mixing matrix structure corresponds to the partition that would be obtained by applying a graph min-cut algorithm to a relational graph constructed from accumulated interaction episodes with the same set of contexts.

23. The system of claim 18, wherein the mixing matrix admits decomposition as a convex combination of permutation matrices by the Birkhoff-von Neumann theorem, providing geometric interpretability wherein the cross-context coherence transfer at any processing cycle is decomposable into a weighted blend of discrete context-swapping operations with non-negative weights.

24. A dual-process cognitive architecture for an autonomous social robot, the architecture comprising: a reflexive processing unit configured to execute real-time sensor processing, context detection, coherence accumulation, application of a cached mixing matrix for manifold-constrained cross-context coherence transfer, behavioral gating, and behavioral output at each processing tick of the robot; a deliberative processing unit configured to execute persistent memory management, context boundary discovery, mixing matrix parameter optimization, and experience-based model updating asynchronously with respect to the reflexive processing unit; and bidirectional modulation pathways connecting the reflexive and deliberative processing units, comprising a downward pathway through which the deliberative processing unit transmits learned modulation data including updated mixing matrix parameters to the reflexive processing unit, and an upward pathway through which the reflexive processing unit transmits interaction data, mixing matrix divergence signals, and conflict

signals to the deliberative processing unit; wherein behavior of the robot emerges from the continuous bidirectional interaction between the reflexive and deliberative processing units rather than from either unit operating in isolation.

25. The architecture of claim 24, wherein the deliberative processing unit generates context-specific stimulus suppression maps from accumulated interaction episodes, each suppression map associating a stimulus pattern with a context key and a suppression factor, and transmits the suppression maps to the reflexive processing unit via the downward pathway, and wherein the reflexive processing unit applies the suppression maps to attenuate reflexive responses to stimuli that have been learned as benign in the current context, such that the same stimulus produces a full reflexive response in an unfamiliar context and an attenuated response in a familiar context where it has been repeatedly observed without negative consequence.

26. The architecture of claim 25, wherein stimulus suppression maps are partially transferable between contexts connected by high-weight entries in the doubly stochastic mixing matrix, such that a suppression rule established in a first context is applied with a reduced suppression factor in a second context that shares substantial coherence with the first context as indicated by the mixing matrix, providing principled partial generalization of learned inhibition proportional to the learned cross-context relationship.

27. The architecture of claim 24, wherein the deliberative processing unit identifies behavioral sequences that have been executed a number of times exceeding a compilation threshold, compresses each identified sequence into a compiled routine comprising a trigger condition, a compressed output sequence, and a distribution signature characterizing sensor conditions during successful executions, and transmits the compiled routines to the reflexive processing unit via the downward pathway, and wherein the reflexive processing unit autonomously executes compiled routines when current conditions match the trigger condition, and halts autonomous execution and returns control to the deliberative processing unit when current sensor conditions deviate from the distribution signature beyond a configurable threshold.

28. The architecture of claim 24, wherein during idle periods the deliberative processing unit performs offline consolidation comprising: replaying accumulated interaction episodes, recomputing trajectory embeddings and relational graph partitions, optimizing mixing matrix parameters and projecting the optimized parameters onto the Birkhoff polytope, cross-validating the mixing matrix structure against the min-cut partition as a boundary quality check, updating stimulus suppression maps, and reviewing compiled routine performance, and transmitting revised coherence accumulators, mixing matrix parameters, context boundaries, suppression maps, and compiled routines to the reflexive processing unit, whereby the robot's

relational understanding changes between active interaction periods as a result of reorganization and reinterpretation of existing experience.

29. The architecture of claim 24, wherein when the reflexive processing unit and the deliberative processing unit produce divergent behavioral assessments for the current context and sensor state, the system enters a conflict resolution state producing an observable behavioral trace comprising a visible hesitation and reduction in behavioral amplitude across output modalities, followed by resolution determined by a context-sensitive priority function, and wherein the conflict event is recorded as an interaction episode informing future deliberative processing.

30. The architecture of claim 29, wherein the context-sensitive priority function resolves conflicts based on at least: the magnitude of divergence between the reflexive and deliberative assessments, the accumulated coherence depth of the current context wherein deeper familiarity increases the weight of the deliberative assessment, and the tension urgency of the current moment wherein higher tension increases the weight of the reflexive assessment.

31. The architecture of claim 29, wherein in the manifold-constrained embodiment, the magnitude of conflict is quantified as the distance between the reflexive processing unit's cached mixing matrix state and the deliberative processing unit's current mixing matrix state measured in the metric space of doubly stochastic matrices, and wherein the hesitation behavior is proportional to the measured distance.

32. The architecture of claim 24, wherein in a degraded mode when the deliberative processing unit is unavailable, the reflexive processing unit continues to operate using locally cached coherence accumulators, stimulus suppression maps, compiled routines, and mixing matrix parameters, with effective coherence computed from instantaneous sensor-derived coherence and the cached mixing matrix, and wherein upon reconnection of the deliberative processing unit, interaction episodes accumulated during the degraded period are transmitted to the deliberative unit for integration into persistent storage, re-optimization of mixing matrix parameters, and recomputation of context boundaries.

33. The architecture of claim 24, wherein the deliberative processing unit computes compressed contextual predictions from accumulated experience and transmits the predictions to the reflexive processing unit as prior expectations, and wherein the reflexive processing unit uses the prior expectations for anticipatory processing, producing faster behavioral response when predictions are confirmed and generating a surprise signal on the upward pathway when predictions are violated.

34. A method for developmental evolution of cross-context coherence transfer in an autonomous social robot, the method comprising: initializing a mixing matrix as a near-identity doubly stochastic matrix with a dynamic gating factor set near zero, such that the robot begins operation with minimal cross-context coherence transfer and predominantly independent coherence accumulation per context; accumulating interaction episodes across a plurality of contexts over the operational life of the robot; periodically optimizing mixing matrix parameters based on accumulated interaction data, the optimization comprising adjustment of static bias parameters representing baseline cross-context relationships and dynamic projection parameters representing sensor-dependent modulation of cross-context transfer; projecting optimized parameters onto the Birkhoff polytope via iterative row and column normalization to maintain the doubly stochastic constraint; and transmitting optimized mixing matrix parameters to a real-time processing unit for application at each processing cycle; whereby the robot's cross-context coherence transfer evolves from minimal transfer (shy, context-isolated behavior) through partial transfer (cautious relational generalization) to structured transfer (rich cross-context relationships with clear domain boundaries), with the developmental trajectory emerging from accumulated experience rather than from predetermined personality parameters.

ABSTRACT

A system and method for managing behavioral expressiveness of an autonomous social robot through Contextual Coherence Fields (CCF) operating within a dual-process cognitive architecture with optional manifold-constrained coherence mixing. The system maintains a plurality of context-keyed coherence accumulators, each independently tracking accumulated relational familiarity for a distinct environmental and social context. An effective coherence value, computed as the minimum of instantaneous sensor-derived coherence and accumulated context coherence, gates the robot's behavioral output, enforcing the architectural invariant that the robot cannot express more behavioral richness than its accumulated familiarity with the present context permits. In a further embodiment, coherence accumulators are organized into a coherence vector and cross-context transfer is governed by a doubly stochastic mixing matrix constrained to the Birkhoff polytope via Sinkhorn-Knopp projection, ensuring that coherence transfer is non-expansive (trust cannot be amplified), compositionally stable (long-term behavioral stability is mathematically guaranteed), and geometrically interpretable (transfer decomposes into weighted permutation blends). The mixing matrix comprises static and dynamic components with the dynamic component gated by a factor initialized near zero,

producing architecturally-enforced shy startup behavior that earns increased cross-context transfer through accumulated interaction. A context boundary discovery subsystem applies graph min-cut algorithms to automatically discover natural context boundaries from accumulated experience, with boundaries additionally derivable from near-zero entries of the mixing matrix in the manifold-constrained embodiment. A dual-process cognitive architecture comprising a reflexive processing unit and a deliberative processing unit enables bidirectional modulation: the deliberative unit generates context-specific stimulus suppression maps (learned inhibition), compiles frequently-repeated behavioral sequences into reflexive routines (habit formation), optimizes mixing matrix parameters, and performs offline consolidation of accumulated experience. Classification conflicts between the two processing units produce observable behavioral hesitation that is interpreted by human observers as evidence of internal depth. The effective coherence and concurrent tension define a behavioral phase space with hysteresis-gated transitions producing distinct behavioral profiles from cautious observation to full expressive fluency.