From Neural Control to Cognitive Explanation: Closed-Loop Negative Feedback and Marr's Levels in YIN-CBEHA

Jyrki Suomala, Suvi Sivén
NeuroLab, Laurea University of Applied Sciences, Finland
(jyrki.suomala, suvi.siven)@laurea.fi

Abstract

Many cognitive theories explain behavior using linear input–output models grounded in normatively specified tasks. However, these approaches remain weakly grounded in biological organization and real-time organism–environment interaction. This theoretical analysis paper examines Yin's control-theoretic extension of Perceptual Control Theory (YIN-CBEHA) using Marr's levels of analysis as evaluative criteria. In YIN-CBEHA, behavior is not identified with motor output but with the hierarchical control of perceptual variables through closed-loop negative feedback. We show how this framework satisfies Marr's three-level demands within a single control architecture. Converging evidence from neurophysiology and robotic construction supports the functional sufficiency of hierarchical feedback control for posture, movement, and spatial regulation under real-world constraints. By constraining admissible computational problems through continuous-time feedback, YIN-CBEHA reduces explanatory underdetermination and improves biological plausibility. While its strongest empirical support currently lies at intermediate sensorimotor levels, the framework provides a tractable foundation for extending control-theoretic explanation toward higher cognition.

Keywords: Hierarchical control; Closed-loop negative feedback; Perceptual Control Theory; Marr's levels of analysis; Biological plausibility

Introduction

Many explanations of behavior in cognitive science have traditionally been framed in terms of linear causal input–output models, often operationalized through reverse-engineering strategies (Paul et al., 2023; Schierwagen, 2012) and formal normative approaches (Anderson, 1991; Oaksford & Chater, 1994). While such approaches aim to decompose cognition into tractable components, they remain insufficiently grounded in the biological organization of living systems and in real-time organism–environment interaction (Bechtel, 2011; Buzsáki, 2019; Guest et al., preprint; Smith & Thelen, 2003). Clarifying how formal models relate to biological organization remains a central theoretical challenge for cognitive science.

These limitations motivate renewed attention to control-theoretic frameworks originating in Perceptual Control Theory (PCT) (Powers, 1973a, 1973b; Powers et al., 1960). Using Marr's levels of analysis as a methodological guide (Marr, 1982), this paper examines Henry Yin's control-theoretical extensions of PCT (Yin, 2014, 2020, 2024, 2016, 2017, 2025), together with converging empirical evidence (Barter et al., 2014, 2015; Barter & Yin, 2021; Bartholomew et al., 2016; Kim et al., 2014). We refer to this integrated account as YIN-CBEHA. Within YIN-CBEHA, behavior is not identified with motor outputs or organized patterns of action, but with the ongoing control of perceptual variables relative to internally specified reference values through hierarchically organized closed-loop feedback (cf. Krakauer et al., 2017; Shadmehr & Krakauer, 2008).

We analyze how YIN-CBEHA meets explanatory demands across Marr's computational, algorithmic, and implementational levels, and contrast its commitments with conventional cognitive modeling approaches. We then evaluate the conditions under which this control-based architecture supports a more general framework for explaining human cognition.

Marr's Levels of Analysis in Cognitive Science

David Marr (1982) introduced his three-level framework to clarify what it means to explain complex information-processing systems. The computational level specifies the problem a system solves and why it is solved, and the algorithmic level characterizes the representations and processes by which the problem is solved, and the implementational level concerns how these processes are physically realized in biological or artificial substrates. Marr emphasized that explanation requires articulation across conceptually distinct levels that function not as competing accounts but as complementary constraints on what counts as a satisfactory theory of cognition.

Marr's framework remains a central evaluative standard in cognitive science, emphasizing that satisfactory explanation requires coherent articulation of computational, algorithmic, and implementational commitments grounded in biological organization and real-time organism–environment interaction (Bechtel, 2011; Krakauer et al., 2017; Marr, 1982; Love, 2015; Webb et al., 2022).

At the same time, interpretations that treat these levels as largely independent have been criticized for insufficient sensitivity to biological organization and real-time organism–environment interaction, particularly in complex living systems (Bechtel, 2011; Krakauer et al., 2017; Marr, 1982).

Against this background, we adopt Marr's three levels as evaluative criteria for analyzing the YIN-CBEHA model (Yin, 2014, 2020, 2024, 2016, 2017, 2025), treating them not as independent descriptive layers but as constraints on explanatory scope and coherence.

We therefore turn next to YIN-CBEHA and its hierarchical organization, proceeding from its core control principles and tracing how increasingly complex forms of behavioral organization emerge. This strategy is loosely analogous to

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Turing's proposal that intelligence should be approached through simple foundational capacities rather than fully developed cognition (Turing, 1950), and it enables a systematic evaluation of YIN-CBEHA's computational commitments, algorithmic organization, and implementational grounding in relation to Marr's explanatory demands.

Neurophysiological grounding of the five lowest levels of hierarchical control

PCT (Powers, 1973a, 1973b; Powers et al., 1960) conceives behavior as emerging from a hierarchy of closed-loop control systems, each regulating a perceptual variable, including both exteroceptive and interoceptive sensory inputs, around a reference value through negative feedback. Control is exerted over sensory input rather than motor output, yielding a circular causal organization linking neural activity, the environment, and perception (Kennaway, 2020, 2025).

While originally proposed as a mathematical model of purposeful behavior without explicit neurophysiological grounding (Powers, 1973b; Powers et al., 1960), recent work within the YIN-CBEHA framework provides convergent evidence that such hierarchical control is implemented in mammalian nervous systems. A mathematical formulation is available in the original sources. Here, we adopt a framework-based theoretical analysis grounded in Marr's levels of analysis, focusing on identifying relevant components and their operations rather than formal modeling, and arguing that explanatory adequacy can be achieved when components, activities, and organizational features are explicitly characterized (Bechtel, 2011).

Empirical findings primarily concern control loops corresponding to the third through fifth hierarchical levels. The two lowest-level loops are muscle tension and muscle length. These are not experimentally isolated in Yin's empirical studies, but are explicitly identified in his theoretical analyses as the foundational spinal control loops upon which higher levels depend (Yin, 2014, 2024, 2025). These loops are well established in classical spinal physiology, including studies of muscle spindles, Golgi tendon organs (GTO), spinal reflex organization, and neuromechanical coupling (Grillner et al., 2008; Houk & Henneman, 1967; McMahon, 2020; Powers, 1973a). When interpreted within a control-theoretic framework, these findings are naturally understood as closed-loop systems rather than stimulus–response pathways (Powers, 1973a; Yin, 2024, 2016). Together, these findings support a five-level hierarchy extending from muscle force to spatial position.

Across the hierarchy, control is realized through closed-loop negative feedback, with lower levels operating faster time scales (Kennaway, 2020, 2025; Marken et al., 2022; Powers, 1973b; Yin, 2014, 2025) and higher levels regulating increasingly abstract perceptual variables (Powers, 1973a; Yin, 2014, 2025).

Although Powers originally proposed a richer hierarchy of up to twelve levels, he emphasized that the number and organization of levels is an empirical matter (Powers, 1973a, 1973b). Higher-order control processes, often associated with cortico–basal ganglia networks, are assumed to regulate increasingly abstract perceptual variables. While these levels currently lack direct neurophysiological grounding, their functional role has been extensively examined behaviorally (Higginson & Mansell, 2008; Marken & Mansell, 2013).

Within the third through fifth levels, basal ganglia output plays a central role in modulating reference signals, contributing to reference specification at multiple hierarchical levels depending on task demands (Yin, 2014, 2016, 2025). In the following section, the five lowest control loops are examined as an integrated hierarchy, proceeding from spinal sensorimotor control to higher levels regulating spatial position.

Levels 1-2. Spinal Control of Force and Length

The lowest level of the hierarchy controls muscle tension (force) and corresponds to Powers' (1973a) intensity level. It constitutes the final common neural pathway through which higher-level control influences the physical world by specifying reference signals for force generation (Yin, 2025).

Muscle tension is sensed by Ib afferents from GTOs, which provide the perceptual input to the loop. Descending supraspinal signals specify a reference signal for desired tension. Perceptual input and reference are compared at the α-motor neuron (α-MN), which functions as the physiological comparator of the loop. Integration of these signals at the α-MN membrane produces an error-dependent firing rate that constitutes the output of the muscle tension loop and drives extrafusal muscle fibers (Grillner et al., 2008; Houk & Henneman, 1967).

Generated muscle force increases tendon tension, thereby modulating Ib-afferent firing from the GTOs. Increased force enhances inhibitory feedback to the α-MN, whereas decreased force reduces it, closing the loop through negative feedback and stabilizing muscle tension around the reference value (Yin, 2025).

The second loop corresponds to Powers' (1973a, 1973b) sensation level. The controlled perceptual variable is muscle length, sensed via Ia and II afferents originating in muscle spindles. The loop operates hierarchically by regulating muscle length around a reference value, while the muscle tension loop provides the force required to achieve this control (Powers, 1973a; Yin, 2014, 2024).

Muscle spindles, embedded in parallel with extrafusial fibers, provide perceptual input by responding to changes in muscle length. Descending γ-motor neuron activity sets the reference for desired muscle length by regulating spindle sensitivity (Yin, 2014, 2024), such that spindle afferent firing reflects deviations from the reference-specified length.

Comparison between reference and perceptual input is implemented through synaptic integration in spinal circuitry, functioning as the physiological comparator. Length-related error signals are transformed via α-MN into a descending reference signal for the muscle tension loop (Houk &

Henneman, 1967; Yin, 2014, 2024), thereby modulating force generation to correct length deviations.

The output of the muscle length loop is a descending reference specifying the force needed to maintain the desired muscle length. The tension loop converts this reference into muscle force, which acts on the musculoskeletal system and intrafusal fibers, altering spindle activity and reducing length error (Yin, 2024). This spindle–musculoskeletal feedback implements negative feedback and stabilizes muscle length around the reference value.

Because the loop is realized largely within spinal circuitry, it can rapidly compensate for disturbances such as gravity, impacts, and limb accelerations. Within this framework, the classical stretch reflex is understood as a closed hierarchical control loop that maintains muscle length according to internally specified reference values (Houk & Henneman, 1967; Yin, 2014, 2024, 2024).

Together, the muscle tension and muscle length loops form a cascaded spinal control system that supports rapid compensation for high-frequency disturbances. In this organization, muscle length control operates at the higher of the two levels by setting reference values for force generation in the muscle tension loop.

Level 3. Joint angle loop

Corresponding to Powers's (1973a) configuration level, the third level of the hierarchy controls joint angle as a geometric relational variable, integrating multiple agonist–antagonist muscle groups into a single perceptual signal (McMahon, 2020; Yin, 2016). This is the first level at which control is explicitly multi-muscular and geometric rather than muscle-specific. Because the same joint configuration can be realized through multiple combinations of muscle lengths and tensions, joint angle cannot be reduced to muscle length (Powers, 1973a; Yin, 2014, 2016).

The perceptual input to the joint angle loop arises from integrated proprioceptive signals from muscle spindles, joint capsule receptors, and cutaneous afferents, forming a unified joint-level representation of current configuration (McMahon, 2020). Descending signals from higher-level loops specify a reference joint configuration, which is compared with perceptual input through distributed synaptic integration within convergent proprioceptive populations, yielding an error signal that specifies required changes in joint configuration (Yin, 2014, 2024, 2016).

The output of the joint angle loop consists of coordinated reference signals to multiple muscle length loops spanning the joint, transforming configuration error into patterns of desired muscle lengths while leaving force and length regulation to lower levels (Yin, 2014, 2024, 2016).

Negative feedback is closed through the mechanical and proprioceptive effects of muscle-length changes on joint geometry, allowing disturbances to be compensated via lower-level control (Yin, 2014, 2016).

Empirical evidence indicates that neuronal firing in striatal and nigral populations covaries continuously with joint-level kinematic variables, including joint angle and movement velocity, during spontaneous locomotion, posture adjustment, and ongoing movement in unrestrained animals (Bartholomew et al., 2016; Kim et al., 2014).

Level 4. Postural orientation loop

Corresponding to Powers's (1973a) transition level, the fourth level of the hierarchy controls whole-body posture and orientation, defined as the global configuration of the body relative to gravity and the support surface. At this level, multiple joints and body segments are integrated into a single perceptual variable that provides a stable spatial and mechanical reference for ongoing behavior. This level operates immediately above the joint angle loop and regulates transitions in global body configuration rather than static joint geometry (Powers, 1973a; Yin, 2020, 2016, 2017, 2025).

The perceptual input to the postural orientation loop arises from multisensory integration of vestibular, proprioceptive, tactile, and visual signals. Together, these signals form a unified postural percept representing global body orientation and stability, rather than the state of any single joint or muscle (McMahon, 2020; Peterka, 2002).

Descending signals from higher-level loops provide a reference signal specifying the desired global body posture and orientation. The comparison between this reference posture and the integrated postural percept is implemented through distributed neural dynamics and their downstream brainstem and midbrain targets. The resulting postural error signal specifies how global body configuration must change to reduce deviation from the reference, while leaving joint- and muscle-level implementation to lower levels of the hierarchy (Yin, 2016, 2017).

The output of the postural orientation loop consists of coordinated reference signals delivered to lower-level joint angle loops, reshaping joint-level reference values to restore or maintain the desired global posture (Yin, 2016, 2017).

Negative feedback at the postural orientation level is closed through the physical interaction between whole-body configuration and the environment, defined primarily by gravity and the support surface. Postural deviations alter multisensory input, generating error signals that adjust descending references and restore global body orientation (Peterka, 2002; Yin, 2016, 2017).

Empirical studies show that neural activity covaries continuously with global body position and orientation during freely moving behavior and typically precedes postural adjustments, consistent with the interpretation that neural signals at this level specify reference values for global body configuration (Barter et al., 2014, 2015; Bartholomew et al., 2016; Kim et al., 2014).

Level 5. Spatial position loop

Corresponding to Powers's (1973a) relationship level, the fifth level of the hierarchy controls spatial position, defined as the organism's location relative to its environment. At this level, the controlled perceptual variables are relational and geometric, regulating where the body is located rather than

its orientation or the movements used to reach it (Bartholomew et al., 2016; Yin, 2014, 2025).

The perceptual input to the spatial position loop arises from the integration of visual, vestibular, and proprioceptive signals into a unified estimate of location relative to environmental features (Barter et al., 2015; Kim et al., 2014; McMahon, 2020).

Descending signals from higher-level loops specify a desired spatial location, and comparison with perceived position yields a continuously varying spatial error signal, such as distance or directional displacement from the target (Barter et al., 2014; Yin, 2016, 2025).

The output of the spatial position loop consists of reference signals delivered to the postural orientation loop, biasing lower-level control so that posture and joint configurations change to reduce spatial error (Yin, 2016).

Negative feedback at the spatial position level is closed through movement-induced changes in multisensory input as the reference location is approached, with disturbances arising primarily from environmental geometry and constraints rather than mechanical perturbations of the body (Barter et al., 2015; Bartholomew et al., 2016; Yin, 2016).

Neural activity recorded during freely moving behavior covaries smoothly with two-dimensional spatial position, independent of specific movements or trajectories, supporting a hierarchical closed-loop control interpretation (Barter et al., 2015; Kim et al., 2014; Yin, 2014, 2016, 2025).

Construction as explanation: robotic validation of hierarchical control

Beyond neurophysiology, YIN-CBEHA can be evaluated through construction by implementing the theory in a physical robot. This reflects a construction-based view of explanation in which theories must be precise enough to generate behavior under real-world constraints (Eliasmith, 2013). Robotics provides a demanding testbed because physical systems must cope with constraints absent from simulations (Chukwurah et al., 2024).

Recent robotic implementations inspired by YIN-CBEHA demonstrate that cascaded negative feedback alone can generate stable and adaptive behavior, without internal models or higher-level cognitive machinery (Barter & Yin, 2021; Johnson et al., 2020; Keino et al., 2026; Yin, 2025). Rather than mirroring biological components one-to-one, these implementations instantiate the core principle of hierarchically organized control of perceptual variables through nested negative feedback. Despite differences in physical substrate, the functional organization is equivalent, making construction an independent line of evidence that complements neurophysiological findings.

Hierarchical control in the robot

The quadruped robot described by Barter and Yin (2021) implements a four-level control hierarchy in which each level regulates its own perceptual variable by setting reference values for the level below. Lower levels operate at faster time scales and interact directly with the mechanical body, whereas higher levels operate more slowly and regulate increasingly abstract variables. Although the number and labels of levels differ from the five neurophysiological levels discussed earlier, the control logic is identical.

At the lowest level, motor torque is regulated via negative feedback from joint sensors. Above this, joint kinematics such as angle and angular velocity are controlled, while higher levels regulate posture and locomotor coordination. The highest level specifies spatial goal variables such as distance or direction to a target, with behavior at all levels emerging through continuous error correction.

Robotic analogue of Level 3: joint angle loop

A correspondence between robotics and neurophysiology can be identified at the joint angle loop, which parallels level 3 in the biological hierarchy. In the robot, joint configuration serves as the controlled perceptual variable. Signals from joint position sensors are integrated into a single representation of joint state and continuously compared with a reference from higher-level controllers, producing an error signal that drives output to lower-level torque controllers and closes the loop (Barter & Yin, 2021).

Disturbances are rejected automatically through negative feedback embedded in the robot's mechanical dynamics. External forces perturb joint input, generate error, and drive compensatory torque (Barter & Yin, 2021). This organization mirrors neurophysiological findings that striatal and nigral activity covaries with joint-level kinematics, illustrating the general principle of controlling perceptual variables through cascaded negative feedback and demonstrating functional sufficiency in a physical system.

The robotic model supports YIN-CBEHA within a construction-based view of explanation by demonstrating the functional sufficiency of hierarchical negative feedback control for generating stable locomotion, posture regulation, and disturbance rejection under real-world constraints. By successfully generating adaptive behavior in a physical system, the robotic implementation provides convergent evidence for the internal coherence and explanatory adequacy of the YIN-CBEHA control architecture. Robotic construction, therefore, complements rather than substitutes for neurophysiological evidence (Keino et al., 2026).

Taken together, the convergence of neurophysiological findings and robotic construction strengthens confidence in YIN-CBEHA as a unified, empirically grounded framework for understanding adaptive behavior across biological and artificial systems.

How Marr's levels appear in YIN-CBEHA: similarities and differences

We use Marr's three levels as evaluative criteria for YIN-CBEHA, focusing on the computational problem, the algorithmic organization, and the biological and physical implementation. Following Krakauer et al. (2017), explanation begins with a functional characterization of behavior prior to interpretation at the level of neural implementation. This section shows how Marr's levels are

jointly realized within the hierarchical control architecture of YIN-CBEHA.

YIN-CBEHA departs from standard formulations in how behavior is defined. Rather than treating behavior as known organized patterns of action or movement execution (Krakauer et al., 2017; Shadmehr & Krakauer, 2008), it conceptualizes behavior as the ongoing outcome of hierarchically organized closed-loop negative feedback control.

Accordingly, explanatory progress depends on identifying which perceptual variables are controlled and maintained despite disturbances. The closed-loop control system serves as the basic explanatory unit, comprising perceptual input, comparison with a reference signal, and output that reduces error within a hierarchy in which higher levels set references for lower levels.

Instead of assigning Marr's computational, algorithmic, and implementational commitments to separate explanatory strata, YIN-CBEHA embeds all three within each control loop. On our interpretation, Marr's computational level need not denote an externally imposed sociocultural metalanguage, but the task definition internal to each control loop: identifying the perceptual variable to be controlled and the disturbance it must resist. In YIN-CBEHA, this 'what/why' specification is realized repeatedly across the hierarchy, since each level defines its own controlled variable, reference condition, and error-correction objective. Empirical inquiry can therefore begin from behavioral, neurophysiological, or robotic evidence while converging on the question of which perceptual variable is being controlled. In this sense, Marr's levels function not as independent descriptive layers but as complementary constraints on the same control organization. Explanatory adequacy emerges when these lines of evidence jointly specify the relevant control loops.

Computational explanation, Marr's levels, and underdetermination

Marr's framework requires that a cognitive theory specify the functional problem an organism solves before proposing implementations. A key challenge for these approaches is that empirical data often underdetermine such problem formulations, as the same observations may support multiple incompatible accounts (Quine, 1951; Van Rooij, 2008). This difficulty is amplified by the fact that distinct internal organizations can generate identical behavior.

Some researchers argue that this underdetermination is not only epistemic but computational, since many computational-level theories permit algorithmic realizations that are biologically intractable (Van Rooij, 2008; Van Rooij & Wareham, 2012). Specifying the computational problem alone often fails to constrain admissible explanations, as the same formulation remains compatible with multiple, radically different algorithms, including algorithms that require global optimization, exhaustive search, or unbounded inference that cannot be implemented by real neural systems (Van Rooij, 2008; Van Rooij & Wareham, 2012).

According to YIN-CBEHA, observable actions are implemented through shared lower-level control mechanisms, whereas behavioral variation reflects how higher-level systems configure reference values. The functional meaning of an action thus derives from its role within a control hierarchy rather than from motor execution itself.

This perspective reframes underdetermination. Rather than asking which algorithm produced a given action, YIN-CBEHA asks which perceptual variables were being controlled, and how their reference values were set within the hierarchy at that moment. This reframing connects the computational, algorithmic, and implementation levels through control-theoretic constraints rather than treating them as separable commitments.

Level 3 as an illustrative case

Level 3 provides a concrete illustration of how Marr's three levels are simultaneously satisfied within a single control loop. The third level of YIN-CBEHA illustrates how Marr's explanatory criteria can be satisfied without introducing ad hoc constructs between Marr's levels (cf. Krakauer et al., 2017). Here, behavior is organized around the control of joint-level configuration variables.

Computational level At this level, the problem is to maintain a configuration-related sensory variable, such as joint configuration or limb posture, within reference limits despite ongoing disturbances (Bartholomew et al., 2016; Kim et al., 2014; Powers, 1973a). In Marr's sense, this specifies what is to be achieved, independently of how it is realized. What must be controlled is a relational sensory variable defined over the organism–environment system. Neurophysiologically, this formulation is consistent with evidence that striatal and nigral neuronal activity covaries with joint-level kinematic variables during ongoing behavior (Bartholomew et al., 2016; Kim et al., 2014).

Algorithmic level At this level, the problem is solved by a negative feedback control loop embedded in a hierarchy (Powers, 1973a; Yin, 2014, 2016). Sensory input is continuously compared against a reference signal supplied by higher levels, and the resulting error drives output that influences the controlled variable through the body and environment (Powers, 1973a). This specifies the algorithmic organization by which the computational goal is achieved. Stability and flexibility arise from the feedback dynamics themselves rather than from auxiliary coordinating constructs, consistent with continuous covariation between neural activity and joint-level kinematics (Bartholomew et al., 2016; Kim et al., 2014).

Implementational level At this level, YIN-CBEHA makes comparatively strong and testable commitments. For loops 3–5, converging neurophysiological evidence implicates cortico-basal ganglia and sensorimotor circuits well suited to continuous-time reference setting and error modulation

(Bartholomew et al., 2016; Kim et al., 2014; Yin, 2014). Complementary robotic implementations further demonstrate that the same hierarchical feedback organization suffices for stable posture and locomotion under real-world disturbances. Together, these findings constrain how the algorithmic organization can be physically realized without collapsing explanation into implementation.

Generalization across levels The same explanatory pattern extends across the hierarchy. At each level, YIN-CBEHA identifies a controlled perceptual variable at the computational level, a negative feedback organization at the algorithmic level, and biologically or physically grounded mechanisms at the implementational level. While empirical constraints are strongest for levels 3–5, the explanatory strategy remains constant, supporting the claim that Marr's criteria are satisfied through repetition of the same control-theoretic schema rather than through level-specific constructs (Yin, 2014, 2016, 2017).

Complexity, tractability, and scope

A recurring challenge for theories of cognition defined at the level of normative task specification concerns whether problems formulated at Marr's computational level admit algorithmic solutions that can realistically be implemented in biological systems. Many influential theories posit problems such as global optimization, exhaustive search, or unconstrained inference whose realizations are computationally intractable under standard complexity assumptions for biological systems (Van Rooij, 2008; Van Rooij & Wareham, 2012). This difficulty arises at the level of problem formulation.

From this perspective, tractability becomes a criterion of explanatory adequacy rather than a mere implementational constraint. YIN-CBEHA addresses this challenge by constraining the space of admissible computational problems through continuous-time closed-loop control (Powers, 1973a; Yin, 2014, 2016). By specifying controlled perceptual variables and feedback organization, the model restricts admissible realizations, increases empirical testability, and makes data more informative for evaluating competing accounts.

Because behavior is governed by real-time negative feedback, computational demands need not scale combinatorially with the number of possible actions or states**.** Control-loop dynamics allow the account to avoid classical intractability results not through approximation, but by redefining the class of problems being solved (Powers, 1973a; Van Rooij, 2008; Yin, 2014, 2016).

YIN-CBEHA does not claim unlimited scope. Its explanatory leverage is strongest at Levels 3–5, where controlled variables, reference signals, and feedback pathways can be empirically constrained through converging behavioral, neurophysiological, and robotic evidence (Bartholomew et al., 2016; Kim et al., 2014; Yin, 2014, 2016). At higher levels of cognition, identifying controlled variables and their reference structures remains an open empirical challenge. Here, YIN-CBEHA functions less as a completed neural account than as a principled framework for formulating tractable questions and guiding empirical inquiry.

This explicit delimitation is a strength rather than a limitation. By tying tractability to hierarchical closed-loop organization, YIN-CBEHA clarifies where strong explanatory claims are warranted and where further development is required. Rather than replacing Marr's framework, the model provides a principled instantiation that renders computational, algorithmic, and implementational claims empirically accountable.

Discussion

This paper examines YIN-CBEHA using Marr's levels as an evaluative criteria that any adequate account of cognition must satisfy. By integrating control-theoretic principles with converging neurophysiological and robotic evidence, the analysis has shown how these demands can be jointly met within a single hierarchical control architecture.

From this perspective, YIN-CBEHA is best understood as a biologically and physically constrained instantiation of Marr's explanatory criteria, grounded in hierarchically organized closed-loop negative feedback control rather than linear causal input–output models. By constraining admissible algorithms through continuous-time feedback and reference-based control, the framework reduces explanatory underdetermination and narrows the space of biologically plausible realizations.

This perspective also resonates with Bernstein's (1935) insight that consistent behavior cannot be explained by fixed motor commands alone, since the same final common path output may yield different outcomes under changing bodily and environmental conditions.

At the same time, YIN-CBEHA makes explicit both its explanatory strengths and its current limits, offering a biologically grounded framework while clarifying where further empirical work is needed.

This paper provides a theory-driven, empirically grounded synthesis but does not offer a full computational specification of YIN-CBEHA. Instead, it presents a theoretical analysis integrating empirical evidence with a qualitative account across Marr's levels. Formal mathematical and computational details are available in the sources (e.g., Powers, 1973a; Kennaway, 2020; Marken et al., 2022; Yin, 2017). Their omission here reflects scope rather than a lack of computational grounding.

Taken together, this situates YIN-CBEHA as a theory of behavioral organization with tractable commitments, explicit limitations, and a principled integration of computational, algorithmic, and implementational considerations grounded in empirical evidence. Rather than replacing existing frameworks, YIN-CBEHA provides a biologically anchored instantiation of Marr's explanatory ideals that clarifies what can currently be explained and where further empirical work is required.

Acknowledgments

The authors used a large language model–based AI tool to assist with language editing, structural clarification, and stylistic refinement of the manuscript. The AI tool was not used to generate original theoretical content, empirical claims, or references. All conceptual contributions, interpretations, and responsibility for the content of the paper remain with the authors.

References

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–485. https://doi.org/10.1017/S0140525X00070801

Barter, J. W., Castro, S., Sukharnikova, T., Rossi, M. A., & Yin, H. H. (2014). The role of the substantia nigra in posture control. *European Journal of Neuroscience*, *39*(9), 1465–1473. https://doi.org/10.1111/ejn.12540

Barter, J. W., Li, S., Lu, D., Bartholomew, R. A., Rossi, M. A., Shoemaker, C. T., Salas-Meza, D., Gaidis, E., & Yin, H. H. (2015). Beyond reward prediction errors: The role of dopamine in movement kinematics. *Frontiers in Integrative Neuroscience*, *9*. https://doi.org/10.3389/fnint.2015.00039

Barter, J. W., & Yin, H. H. (2021). Achieving natural behavior in a robot using neurally inspired hierarchical perceptual control. *IScience*, *24*(9), 102948. https://doi.org/10.1016/j.isci.2021.102948

Bartholomew, R. A., Li, H., Gaidis, E. J., Stackmann, M., Shoemaker, C. T., Rossi, M. A., & Yin, H. H. (2016). Striatonigral control of movement velocity in mice. *European Journal of Neuroscience*, *43*(8), 1097–1110. https://doi.org/10.1111/ejn.13187

Bechtel, W. (2011). Mechanism and Biological Explanation. *Philosophy of Science*, *78*(4), 533–557. https://doi.org/10.1086/661513

Bernstein, N. A. (1935). The problem of the interrelation of coordination and localization. *Archives of Biological Sciences*, *38*, 15–59.

Buzsáki, G. (2019). *The Brain from Inside Out* (1st ed.). Oxford University PressNew York. https://doi.org/10.1093/oso/9780190905385.001.0001

Chukwurah, N., Adebayo, A. S., & Ajayi, O. O. (2024). Sim-to-Real Transfer in Robotics: Addressing the Gap between Simulation and Real-World Performance. *Journal of Frontiers in Multidisciplinary Research*, *5*(1), 33–39. https://doi.org/10.54660/.IJFMR.2024.5.1.33-39

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.

Grillner, S., Wallén, P., Saitoh, K., Kozlov, A., & Robertson, B. (2008). Neural bases of goal-directed locomotion in vertebrates—An overview. *Brain Research Reviews*, *57*(1), 2–12. https://doi.org/10.1016/j.brainresrev.2007.06.027

Guest, O., Scharfenberg, G., & Van Rooij, I. (preprint). Modern alchemy: Neurocognitive reverse engineering. *Preprint*.

Higginson, Sally., & Mansell, Warren. (2008). What is the mechanism of psychological change? A qualitative analysis of six individuals who experienced personal change and recovery. *Psychology and Psychotherapy: Theory, Research and Practice*, *81*(3), 309–328. https://doi.org/10.1348/147608308X320125

Houk, J., & Henneman, E. (1967). Responses of Golgi tendon organs to active contractions of the soleus muscle of the cat. *Journal of Neurophysiology*, *30*(3), 466–481. https://doi.org/10.1152/jn.1967.30.3.466

Johnson, T., Zhou, S., Cheah, W., Mansell, W., Young, R., & Watson, S. (2020). Implementation of a Perceptual Controller for an Inverted Pendulum Robot. *Journal of Intelligent & Robotic Systems*, *99*(3–4), 683–692. https://doi.org/10.1007/s10846-020-01158-4

Keino, L., Sivén, S., & Suomala, J. (2026). *An analog robot that resembles the human mind*. Laurea Journal. http://urn.fi/URN:NBN:fi-fe2026041628017

Kennaway, R. (2020). When causation does not imply correlation. In *The Interdisciplinary Handbook of Perceptual Control Theory* (pp. 49–72). Elsevier. https://doi.org/10.1016/B978-0-12-818948-1.00004-6

Kennaway, R. (2025). Perceptual control theory and the free energy principle: A comparison. *Current Opinion in Behavioral Sciences*, *65*, 101575. https://doi.org/10.1016/j.cobeha.2025.101575

Kim, N., Barter, J. W., Sukharnikova, T., & Yin, H. H. (2014). Striatal firing rate reflects head movement velocity. *European Journal of Neuroscience*, *40*(10), 3481–3490. https://doi.org/10.1111/ejn.12722

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Love, B. C. (2015). The Algorithmic Level Is the Bridge Between Computation and Brain. *Topics in Cognitive Science*, *7*(2), 230–242. https://doi.org/10.1111/tops.12131

Marken, R., Kennaway, R., & Gulrez, T. (2022). Behavioral illusions: The Snark *is* a Boojum. *Theory & Psychology*, *32*(3), 491–514. https://doi.org/10.1177/09593543211070271

Marken, R. S., & Mansell, W. (2013). Perceptual Control as a Unifying Concept in Psychology. *Review of General Psychology*, *17*(2), 190–195. https://doi.org/10.1037/a0032933

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman.

McMahon, T. A. (2020). *Muscles, Reflexes, and Locomotion*. Princeton University Press. https://doi.org/10.2307/j.ctv173f1sp

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological*

Review, *101*(4), 608–631. https://doi.org/10.1037/0033-295X.101.4.608

Paul, L. A., Ullman, T. D., De Freitas, J., & Tenenbaum, J. (2023). *Reverse-engineering the self*. https://doi.org/10.31234/osf.io/vzwrn

Peterka, R. J. (2002). Sensorimotor Integration in Human Postural Control. *Journal of Neurophysiology*, *88*(3), 1097–1118. https://doi.org/10.1152/jn.2002.88.3.1097

Powers, W. T. (1973a). *Behavior: The control of perception*. Aldine Publ.

Powers, W. T. (1973b). Feedback: Beyond Behaviorism: Stimulus-response laws are wholly predictable within a control-system model of behavioral organization. *Science*, *179*(4071), 351–356. https://doi.org/10.1126/science.179.4071.351

Powers, W. T., Clark, R. K., & Farland, R. L. M. (1960). A General Feedback Theory of Human Behavior: Part I. *Perceptual and Motor Skills*, *11*(1), 71–88. https://doi.org/10.2466/pms.1960.11.1.71

Quine, W. V. (1951). Main Trends in Recent Philosophy: Two Dogmas of Empiricism. *The Philosophical Review*, *60*(1), 20. https://doi.org/10.2307/2181906

Schierwagen, A. (2012). *On Reverse Engineering in the Cognitive and Brain Sciences* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1201.4896

Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, *185*(3), 359–381. https://doi.org/10.1007/s00221-008-1280-5

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. https://doi.org/10.1016/S1364-6613(03)00156-6

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460.

Van Rooij, I. (2008). The Tractable Cognition Thesis. *Cognitive Science*, *32*(6), 939–984. https://doi.org/10.1080/03640210801897856

Van Rooij, I., & Wareham, T. (2012). Intractability and approximation of optimization theories of cognition. *Journal of Mathematical Psychology*, *56*(4), 232–247. https://doi.org/10.1016/j.jmp.2012.05.002

Webb, R., Zhao, C. Y., Osborne, M., Landry, P., & Camerer, C. F. (2022). A Neuro-Autopilot Theory of Habit: Evidence from Canned Tuna. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4130496

Yin, H. (2014). How Basal Ganglia Outputs Generate Behavior. *Advances in Neuroscience*, *2014*. http://dx.doi.org/10.1155/2014/768313

Yin, H. (2020). Chapter 3—The crisis in neuroscience. In W. Mansell (Ed.), *The Interdisciplinary Handbook of Perceptual Control Theory* (pp. 23–48). Academic Press. https://doi.org/10.1016/B978-0-12-818948-1.00003-4

Yin, H. (2024). *The integrative functions of the basal ganglia* (First edition). CRC Press.

Yin, H. H. (2016). The Basal Ganglia and Hierarchical Control in Voluntary Behavior. In J.-J. Soghomonian (Ed.), *The Basal Ganglia* (pp. 513–566). Springer International Publishing. https://doi.org/10.1007/978-3-319-42743-0_20

Yin, H. H. (2017). The Basal Ganglia in Action. *The Neuroscientist*, *23*(3), 299–313. https://doi.org/10.1177/1073858416654115

Yin, H. H. (2025). Aligning brain and behavior. *Current Opinion in Behavioral Sciences*, *62*, 101487. https://doi.org/10.1016/j.cobeha.2025.101487