Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Symbolic feedback in social coordination: A cognition-based control-theoretic framework

Ehud Ahissar

ehud.ahissar@weizmann.ac.il

Weizmann Institute of Science

Article

Keywords: collective coordination, brain–brain communication, control theory, symbolic systems, societal regulation, pantheism

Posted Date: October 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10830730/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Human societies maintain coordination among large numbers of cognitively autonomous individuals whose beliefs, expectations, and goals are only locally informed. How shared symbolic systems contribute to such coordination remains a central problem across the social and cognitive sciences. This paper proposes a control-theoretic framework in which symbolic communication creates a population-level feedback loop between individual cognition and a shared symbolic state. Building on the BWBB architecture, which distinguishes analog brain-world interactions from digitally structured brain-brain interactions, the framework models the brain-brain domain as a supra-individual dynamical system. Individuals contribute communicative signals that update shared norms, symbols, and expectations, while this shared state feeds back to regulate their internal cognitive models. Regulation operates through complementary mechanisms analogous to proportional control through overt enforcement, derivative-like control through internalized and anticipatory monitoring, and integral-like control through accumulated social and economic consequences. Population-level coordination is characterised by bounded variation among internal models relative to the shared symbolic state, allowing social coherence without requiring cognitive uniformity. Religious systems are analysed as historically persistent implementations of this architecture, and concepts of divine authority as symbolic representations of the integrative closure of the societal feedback loop. More broadly, the framework proposes that shared symbolic representations do not merely express social organisation but actively participate in regulating it, providing a common theoretical language for analysing religion, institutions, and secular systems of social coordination.

1. Introduction

Human societies persist or collapse depending on their ability to maintain coordination among large numbers of individuals. While individuals are biological organisms, societies function as structured ensembles whose survival depends on sustained coordination across large numbers of agents (Luhmann, 1995; Parsons, 1951). Throughout history, societies have differed dramatically in their stability, longevity, and capacity for collective action (Turchin, 2016). These differences are likely linked to variation in the transmission and regulation of symbolic systems that maintain coordination at the population level (Boyd & Richerson, 2005; Cavalli-Sforza & Feldman, 1981; Henrich, 2015).

The central problem addressed in this paper is how large groups of cognitively autonomous individuals achieve sufficiently stable alignment to function as coherent systems over time. Individual cognition is inherently bounded and locally constrained (Simon, 1957). Each agent operates with partial information, limited predictive capacity, and private goals. Without additional regulatory structure, such systems are vulnerable to collective action failure and coordination breakdown (Axelrod, 1984; Hardin, 1968; Ostrom, 1990).

I propose that large human societies can be understood as symbolic feedback control systems. In this framework, distributed individuals continuously construct a shared symbolic state through

communication, while that shared state feeds back to regulate their internal cognitive models. This reciprocal interaction constitutes a population-level control architecture that enables large-scale social coordination. I further argue that the cognitive substrate implementing this symbolic control architecture is provided by the recently proposed BWBB framework (Ahissar, Polani, & Ahissar, 2025), in which cognition operates within two coupled domains: the BW-domain, governing analog brain–world interaction, and the BB-domain, governing digitally structured brain–brain communication. The relationship between these two domains is illustrated schematically in Fig. 1.

The symbolic control architecture proposed here depends critically on the properties of the BB-domain. Although instantiated in individual brains, the BB-domain is not reducible to any single agent. It emerges from distributed communication among agents and can be understood as a supra-individual dynamical system (List & Pettit, 2011; Theiner, Allen, & Goldstone, 2010) whose state reflects the evolving configuration of shared digital representations across the population. Societal coordination arises from this closed feedback loop: local cognitive processes construct and update the shared representational field, and that field, in turn, feeds back normative and predictive constraints onto individual cognition.

Within this loop, coordination is not achieved solely through overt enforcement or explicit agreements (Young, 1996). Regulation operates through two complementary channels. Overt control involves publicly communicated norms, institutionalized rules, and explicit sanctions (North, 1990; Tyler, 2006). Covert control operates through internalized expectations, anticipatory evaluation, and self-regulation shaped by the shared domain (Carver & Scheier, 1998; Tomasello, 2016).

Religious systems provide a particularly clear instantiation of this architecture (Sosis & Alcorta, 2003). In such systems, the concept of God represents the integrative wholeness of the societal feedback loop: the totality of the BB-domain construed as an authoritative and evaluative unity. God, in this framework, is not treated as a metaphysical postulate but as the representational encoding of the global control structure through which society regulates itself (Durkheim, 1995; Norenzayan, 2013; Whitehouse et al., 2019). By stabilising both overt norms and covert internalized monitoring, religious systems implement a scalable mechanism for maintaining coordination at the level of the social whole.

More generally, this framework contributes to ongoing debates in philosophy of mind and social cognition concerning the nature of supra-individual cognitive systems, the status of collective intentionality, and the role of shared symbolic structures in regulating individual cognition (Clark & Chalmers, 1998; List & Pettit, 2011; Searle, 1995). It departs from existing approaches in several key respects. Searle's analysis of social reality treats institutions as systems of collectively recognized status functions grounded in individual intentional states (Searle, 1995), such that social structures ultimately depend on meaning assignments made by individual minds. By contrast, the BB-domain is modelled here as a supra-individual dynamical system whose regulatory properties cannot be reduced to individual intentionality, even though it is instantiated through it.

Similarly, accounts of group agency characterise collectives as agents with decision-making structures defined over individual inputs (List & Pettit, 2011). Rather than treating the group as the primary locus of

agency, the present approach emphasizes how distributed symbolic representations regulate individual cognition. Finally, extended and distributed cognition frameworks argue that cognitive processes can extend beyond the individual through interaction with external artifacts and other agents (Clark & Chalmers, 1998; Hutchins, 1995). The BB-domain framework builds on this insight but shifts the focus from the extension of cognition to its regulation at the population level through shared symbolic states. This shift reframes supra-individual cognition not as an extension of individual minds, but as a control structure that stabilises coordination across populations.

2. Coordination as a Multi-Level Control Problem

Individual cognition can be understood as a control system operating under bounded informational and computational constraints (Powers, 1973; Wiener, 1948). At the level of the individual, behaviour is regulated through feedback processes in which perceived states are compared to internal reference conditions and actions are adjusted to reduce discrepancies (Carver & Scheier, 1998). Such regulation is necessarily local: each agent maintains internal variables within viable bounds relative to its own goals and predictions. Because cognitive resources are limited and information is incomplete, this regulation is constrained by bounded rationality (Simon, 1957).

Local control, however, does not guarantee collective coherence. Two individuals may each act optimally relative to their own internal reference values while jointly generating instability at the group level. In collective action settings, individually rational choices can fail to produce jointly beneficial coordination (Axelrod, 1984; Fehr & Gächter, 2000). As interaction networks expand and populations scale, informal alignment mechanisms become insufficient: monitoring weakens, incentives to free ride increase, and small divergences in expectations can propagate across agents (Young, 1996). Under such conditions, representational misalignment accumulates, rendering coordination increasingly fragile (Olson, 1965).

Population-level stability therefore requires more than individual self-regulation. It requires convergence among the normative and predictive models held by different agents. When internal expectations diverge sufficiently, coordination becomes unstable even if no individual intends disruption. As group size increases, direct interpersonal monitoring and dyadic enforcement cannot reliably maintain this convergence. Empirical analyses of collective governance show that stable cooperation at scale depends on structured monitoring, sanctioning, and rule systems that extend beyond informal alignment (Ostrom, 1990). A higher-order regulatory structure is therefore required to reduce variance in shared expectations and stabilise alignment across the population. The question then becomes what kind of cognitive architecture can implement such population-level regulation while remaining compatible with the constraints of individual cognition.

3. The BWBB Architecture

The BWBB framework (Ahissar et al., 2025) provides a cognitive architecture capable of implementing the population-level regulation required for large-scale coordination. It distinguishes two domains of

interaction: the BW-domain, which governs analog brain–world coupling, and the BB-domain, which governs digitally structured brain–brain communication. Within this architecture, the BB-domain functions as a supra-individual regulatory layer through which alignment among distributed internal models can be stabilised. This distinction allows coordination to be analysed not only as a social phenomenon but as a consequence of interacting cognitive processes distributed across individuals.

The BW-domain comprises brain–world interactions. In this domain, cognitive systems regulate behaviour through continuous sensorimotor coupling with the physical environment. Representations are grounded in analog processes: graded signals, continuous feedback, and dynamically varying sensory inputs (Ashby, 1956; Wiener, 1948). Control in the BW-domain maintains local stability through convergence in motor–sensory–motor loops, in which sensor motion and sensory feedback are reciprocally integrated until stable perceptual states are achieved (Ahissar & Assa, 2016). Within these loops, predictive coding mechanisms operate as local components that minimize discrepancies between expected and incoming signals (Friston, 2010). Because regulation in the BW-domain depends on continuous, embodied sensorimotor coupling with the environment, it is necessarily individual and analog in structure.

In contrast, the BB-domain comprises brain–brain interactions mediated by symbolically structured communication. Language, ritual, writing, and other combinatorial signalling systems allow agents to exchange discrete tokens that encode categorical distinctions and shared models (Tomasello, 2007). It is therefore digital in representational format. Its elements are not continuous sensorimotor states but discrete, socially recognized symbolic structures. These digital representations can be transmitted, stored, recombined, and stabilised across individuals (Clark & Chalmers, 1998; Searle, 1995).

The digital character of the BB-domain does not imply that neural processes themselves are binary or computational in a simplistic sense. Rather, digitality refers to the structure of shared symbolic representations: they operate through discrete categories, rule-governed combinations, and shared symbolic distinctions. Whereas the BW-domain regulates interaction with the physical world through continuous feedback, the BB-domain regulates interaction among agents through discrete, collectively maintained constraints.

Crucially, the BB-domain is not reducible to the BW-domain. Although symbol production and perception depend on neural processes, the representational structures that constitute the BB-domain exist at the level of coordinated inter-brain dynamics. A linguistic norm, a legal rule, or a shared moral category cannot be localized within a single individual's brain; it is defined by distributed agreement and mutual recognition (Hutchins, 1995; List & Pettit, 2011; Theiner et al., 2010). The BB-domain therefore constitutes a supra-individual representational layer that emerges from but is not contained within any single cognitive system (Clark, 1998; Hutchins, 1995).

Because the BB-domain is digital and population-distributed, it can function as a higher-order control structure, consistent with classical accounts of social systems as self-regulating structures (Luhmann, 1995; Parsons, 1951). Digital symbolic systems allow stable reference values to be encoded explicitly,

transmitted across generations, and reinforced through collective practices (Boyd & Richerson, 2005; Henrich et al., 2010). Shared representations can thus serve as common reference points against which individual cognition is evaluated. In this way, the BB-domain provides the structural conditions necessary for population-level alignment.

The distinction between BW and BB domains clarifies why large-scale coordination cannot be reduced to individual-level regulation alone. Individual BW-level control maintains local stability, but convergence of internal models across agents requires a digitally mediated representational field. It is this field — the BB-domain — that can serve as the supra-individual control layer required for societal persistence. In the next section, I formalize this layer as a dynamical system and specify the feedback relations that couple it to individual cognition.

4. Formal Model of the BB-Domain

Population-level stability requires convergence among distributed internal models, and the BB-domain provides the structural basis for such convergence. The BB-domain can be understood as a supra-individual dynamical system that regulates individual cognition through a population-level digital representational state organised as a closed feedback loop. At the individual level, this regulation is implemented through the updating of internal predictive and evaluative models in response to internally represented symbolic constraints. Figure 2 summarizes the proposed closed-loop architecture before its formal specification.

State Variables

Let the BB-domain at time t be represented by a population-level state S(t). S(t) summarizes the configuration of shared digital representations in a society, including norm categories, rule structures, evaluative standards, and shared predictive schemas. This state is not located in any individual. It is distributed across agents and stabilised through public representation, collective practice, and institutional persistence.

At the individual level, each agent i maintains an internal model $M_i\,(t)$. $M_i\,(t)$ includes that individual's beliefs about norms, expected evaluations, likely sanctions, and the behaviour of others. Population-level coordination depends on the degree of alignment between the set of internal models [$M_i\,(t)$] and the shared symbolic configuration S(t) (Young, 1998). Here and below, square brackets denote the set of variables across all agents in the population.

Closed Loop between Internal Models and BB State

Agents contribute to the BB-domain through digitally structured signalling, including speech, writing, ritual participation, and norm-enforcing actions. Let the BB-relevant contribution of agent i at time t be denoted $I_i\,(t)$. To make the loop explicit, I specify $I_i\,(t)$ as a function of the agent's internal model:

$$I_i\,(t) \;=\; H_i\,(M_i(\,t),\; S(t),\; X(t\,))$$

where X(t) denotes contextual conditions, including environmental pressures and social situations. H_i captures how internal expectations and motivations generate outward symbolic behaviour.

The BB-domain integrates distributed inputs and updates according to:

S(t + 1) = *F*(S(t), [I_i(t)], X(t))

The function *F* captures processes such as reinforcement, conformity pressure, institutional codification, selective amplification, and suppression of destabilizing signals. Because the BB-domain is digital in representational format, state transitions are often categorical rather than continuous. Norms may be affirmed, revised, or repudiated in discrete symbolic steps rather than gradually adjusted along a continuum.

Overt and Covert Feedback

The BB-domain regulates individuals by feeding back constraints that shape internal models. I distinguish two complementary channels.

Overt feedback consists of publicly accessible rules, explicit sanctions, and institutionalized enforcement.

$$O_i\,(\mathrm{t}) \;=\; G_o\,(\mathrm{S}(\,\mathrm{t}),\;\mathrm{X}(\mathrm{t}\,))$$

Covert feedback consists of internalized evaluation, anticipatory monitoring, and self-regulation shaped by the shared domain.

$$C_i\,(\mathrm{t}) \;=\; G_c\,(M_i(\,\mathrm{t}),\;\mathrm{S}(\mathrm{t}),\;\mathrm{X}(\mathrm{t}\,))$$

Individual internal models update according to:

M_i (t + 1) = U(M_i(t), O_i(t), C_i(t), X(t))

In these formulations, the mapping functions (H_i, F, G_o, G_c, U) serve as conceptual operators that formalize the interaction between individual cognitive states and population-level symbolic dynamics. They represent the translation between the continuous, analog processes of the BW-domain and the discrete, digital constraints of the BB-domain. The contextual variable $X\,(t)$ serves as a boundary condition for these interactions, capturing systemic pressures such as population scale, resource constraints, or environmental volatility.

Coordination Error

Population-level coordination error can be defined as:

E(t) = Var([M_i(t)] | S(t))

In this formulation, the conditional bar indicates that variance is evaluated relative to the current BB-domain state S(t). Coordination error therefore reflects the dispersion of individual internal models around the shared symbolic configuration. In computational terms, this conditional variance can be interpreted as the degree of divergence among internal models relative to the shared symbolic configuration. High values of E(t) indicate representational divergence and instability, whereas low values indicate convergence and coordinated alignment. The control function of the BB-domain is to reduce E(t) over time by aligning internal models through shared digital constraints. Empirically, $E(t)$ could be approximated by measuring the dispersion of agents' normative expectations or judgements relative to the publicly articulated symbolic configuration, for example through surveys, behavioural experiments, or analysis of communicative signals that reveal agreement or disagreement about shared norms.

Stability Conditions

Stability at the societal level requires sufficient persistence of the BB-domain state $S(t)$, effective feedback from $S(t)$ to the internal models $M_i(t)$, and bounded divergence among those models. If distributed inputs $I_i(t)$ destabilize $S(t)$ faster than feedback can realign $M_i(t)$, coordination error increases. If feedback is too weak, variance in internal models accumulates and the system drifts away from the shared symbolic configuration. If feedback is excessively rigid, adaptability decreases and the system becomes brittle (Carlson & Doyle, 2002). A stable BB-domain therefore corresponds to a regime in which coordination error $E(t)$ remains bounded despite continual perturbations generated by distributed inputs. A stable society does not require complete convergence of internal models, but rather bounded variance compatible with coordination and adaptive flexibility. The formalization developed here generates experimentally testable predictions concerning the effects of symbolic integration, overt and covert regulation, and coordination error on the stability of large-scale social systems.

5. Religion as a Stabilised BB Control Architecture

The formal model developed in the previous section describes a general mechanism through which a population-level symbolic state $S(t)$ regulates distributed internal models $M_i(t)$ through a closed feedback loop. The key question is how such a control architecture can be stabilised in large human societies. Religious systems provide a historically widespread implementation of this structure.

Religious traditions generate and maintain digital symbolic structures that stabilise the BB-domain (Atran & Henrich, 2010). Norms, narratives, rituals, and canonical texts encode shared representations in discrete symbolic form and transmit them through repeated communication and participation. Because these structures are publicly represented and socially reinforced, they persist across generations and function as stable reference values for evaluating behaviour. In terms of the model introduced in Section 4, religious systems help stabilise the population-level state $S(t)$ by preserving and reinforcing symbolic constraints that guide individual internal models.

Beyond simple persistence, religious symbolic systems often integrate normative expectations, evaluative standards, and mechanisms of enforcement into a unified structure. This integration increases the coherence of the BB-domain and reduces variance among internal models. Empirical research on religion and cooperation shows that shared religious beliefs and practices can increase norm compliance and stabilise cooperation in large groups, particularly when those beliefs include expectations of monitoring and evaluation beyond immediate social observation (Purzycki et al., 2016; Shariff & Norenzayan, 2011).

Overt and Covert Feedback

A central feature of religious regulatory structures is the integration of overt and covert feedback channels. These channels correspond to distinct mechanisms through which the BB-domain constrains individual cognition and behaviour. The correspondence between the proposed social mechanisms and their control-theoretic analogs is summarized in Fig. 3.

Overt feedback

Overt feedback operates through publicly observable rules, institutionalized norms, and explicit sanctioning. Religious communities articulate behavioural constraints and enforce them through collective monitoring and social penalties. Such structures provide visible signals that reinforce the shared symbolic configuration encoded in $S(t)$.

Many of the behavioural prohibitions in the Ten Commandments illustrate overt feedback mechanisms. Commands such as “You shall not murder,” “You shall not steal,” and “You shall not bear false witness” regulate observable actions that can be monitored by other members of the community and corrected through public sanction.

In control-theoretic terms, this mechanism resembles *proportional feedback*. Corrective signals are generated in proportion to deviations from the reference state (Astrom & Murray, 2010; Franklin, Powell, & Emami-Naeini, 2015). When behaviour diverges from the normative configuration encoded in $S(t)$, overt sanctions produce corrective pressure that pushes the system back toward alignment.

Covert feedback

Covert feedback operates through internalized evaluation rather than external enforcement. Through teaching, ritual participation, and symbolic narratives, individuals incorporate the evaluative structure of the BB-domain into their internal models $M_i(t)$. As a result, deviations from shared norms generate anticipatory internal signals even in the absence of external observers.

Several commandments illustrate this mechanism. The prohibition against coveting regulates internal mental states rather than externally observable actions. Other commandments, such as those emphasizing reverence or fidelity to divine authority, function primarily through internalized expectations rather than continuous social monitoring.

In the formal model, these processes correspond to the covert feedback channel $C_i(t)$. Internalized expectations generate anticipatory signals such as guilt, shame, or moral concern when contemplated behaviour diverges from the shared symbolic configuration. These signals guide behaviour before overt deviations occur.

In control-theoretic terms, covert regulation resembles *derivative feedback*, which anticipates changes in system state and generates corrective signals before large deviations develop (Astrom & Murray, 2010). By embedding evaluative standards within internal models, religious systems enable early detection and correction of emerging coordination errors. However, as in engineered systems, derivative control can be highly sensitive to noise (Franklin et al., 2015); an overly reactive covert feedback channel may manifest culturally as systematic anxiety or obsessive rule-following.

Integrated control

The integration of overt and covert feedback allows religious systems to regulate behaviour across large populations despite limited direct monitoring. Overt mechanisms maintain publicly visible normative boundaries, while covert mechanisms allow individuals to regulate themselves when external enforcement is absent. Together these channels reduce coordination error $E(t)$ by aligning distributed internal models with the shared symbolic configuration.

In addition to these proportional and anticipatory components, long-term societal consequences may provide an integral-like feedback signal. Sustained coordination can generate cumulative economic and social benefits, including increased stability, trust, and resource availability, which in turn reinforce adherence to the shared symbolic configuration. Conversely, prolonged coordination failure may accumulate destabilizing pressures that weaken compliance and increase divergence among internal models. In this sense, economic and societal outcomes can function as an accumulated feedback term operating over extended timescales.

This combination of symbolic persistence, overt enforcement, and internalized monitoring allows religious systems to stabilise the BB-domain as a supra-individual coordination system. The resulting symbolic structure not only constrains individual behaviour but also generates a coherent representation of the normative authority underlying the system. The next section examines how this representation is expressed in many religious traditions through the concept of God as the integrative representation of the societal control loop.

6. God as Representational Closure of the Societal Control Loop

The preceding sections described how a population-level symbolic state $S(t)$ regulates distributed internal models $M_i(t)$ through a closed feedback loop mediated by the BB-domain. Religious systems stabilise this architecture by maintaining persistent symbolic structures and by integrating overt and covert feedback mechanisms. A central feature of many religious traditions is the representation of this

regulatory structure in the form of a unified evaluative authority, typically expressed through the concept of God.

Within the framework developed here, the concept of God can be interpreted as a symbolic representation of the integrative wholeness of the BB-domain. As described in the formal model above, the BB-domain functions as a supra-individual control layer that aggregates distributed contributions from agents and feeds back normative constraints onto their internal models. However, this distributed regulatory structure is not directly observable as a single entity. Religious systems encode it symbolically as a unified authority that is perceived as monitoring, judging, and guiding individual behaviour. Classical sociological analyses of religion have long noted that the sacred often represents the collective order of society itself rather than a merely individual belief system (Bellah, 1967; Durkheim, 1995).

This representation performs an important stabilising function. Without a unified symbolic representation, the BB-domain remains a distributed structure whose authority must be inferred from the behaviour and expectations of many agents. The representation of a single evaluative authority compresses this distributed structure into a cognitively tractable form. Individuals can treat the normative order as if it were generated by a coherent and persistent source of evaluation. Research on cultural cognition and cooperation suggests that such representations of supernatural monitoring can enhance norm compliance and cooperation in large groups (Norenzayan, 2013; Purzycki et al., 2016; Shariff & Norenzayan, 2011).

The concept of divine omniscience further reinforces the covert feedback channel described in the previous section. If the symbolic authority is represented as continuously aware of individual behaviour and intentions, internalized monitoring becomes independent of the presence of human observers. Individuals anticipate evaluation even when external monitoring is absent, thereby maintaining alignment between internal models and the shared symbolic configuration. In the formal model, the representation of divine monitoring strengthens the feedback mapping from $S(t)$ to $M_i(t)$ by ensuring that the covert error signal $C_i(t)$ remains active across contexts.

Importantly, the concept of God represents more than a set of rules. It represents the unity and authority of the entire normative system. By encoding the BB-domain as a single coherent source of evaluation, religious systems anchor the symbolic configuration $S(t)$ and reinforce its stability across the population. Individuals can therefore interpret normative constraints as originating from a stable reference point rather than from a fluctuating set of interpersonal expectations.

In control-theoretic terms, the concept of God functions as a representation of loop closure (Ashby, 1956; Von Foerster, 2003) (see Fig. 2). This representation allows individuals to treat distributed social evaluation as a single coherent source of normative authority. In the formal model of the BB-domain, agent contributions are integrated into the shared symbolic state, which in turn feeds back to regulate individual internal models. The symbolic representation of God renders this structure cognitively explicit by portraying the regulatory system as an integrated authority that both receives information about behaviour and generates evaluative feedback. By making the feedback loop perceptually and

conceptually unified, this representation strengthens the perceived coherence of the BB-domain and promotes convergence among internal models.

The stabilising role of this representation does not depend on the metaphysical status of the entity it describes. What matters for the functioning of the control architecture is that individuals treat the symbolic authority as real and authoritative within their internal models. When this condition holds, the representation of God helps maintain the closed feedback loop between distributed agents and the population-level normative state.

Consequences of the Framework

The framework developed here has several consequences for understanding the role of symbolic authority in societal coordination:

- The concept of God represents the integrative wholeness of the BB-domain and functions as a symbolic representation of the population-level control structure.
- Representing the BB-domain as a unified evaluative authority reduces the cognitive complexity of interpreting distributed normative expectations.
- Belief in an omniscient evaluative authority strengthens covert feedback by sustaining anticipatory monitoring even when external observers are absent.
- By reinforcing both overt and covert feedback channels, the representation of God stabilises the closed control loop between individual internal models and the population-level symbolic configuration.

7. Implications

The framework developed in this paper treats societal coordination as the outcome of a supra-individual regulatory system operating through the BB-domain. By modelling the BB-domain as a population-level dynamical system that regulates distributed internal models, the analysis shifts the explanatory focus from individual cognition alone to the interaction between individuals and the shared symbolic structures that constrain them. Several implications follow from this perspective.

First, the framework clarifies why large-scale societies require stable symbolic structures that extend beyond direct interpersonal interaction. As populations grow, dyadic monitoring and informal alignment become insufficient to maintain convergence among internal models. Persistent symbolic configurations in the BB-domain allow societies to stabilise shared expectations across large and heterogeneous populations. Research on cultural transmission and cultural evolution has shown that socially transmitted symbolic systems can accumulate across generations and stabilise large-scale coordination (Boyd & Richerson, 2005; Cavalli-Sforza & Feldman, 1981; Henrich, 2015). Religious systems represent one historically important mechanism for achieving such stabilization.

Second, the analysis suggests that societal stability depends on the balance between persistence and adaptability in the BB-domain. If the shared symbolic configuration $S(t)$ changes too rapidly, individuals cannot update their internal models quickly enough to maintain alignment, and coordination error increases. Conversely, if symbolic structures become too rigid, the system loses the ability to adapt to environmental and social change. Empirical work on institutional governance shows that stable cooperative systems require mechanisms that preserve shared rules while allowing incremental adaptation (Ostrom, 1990). Stable societies therefore operate within a regime in which symbolic constraints are persistent enough to maintain alignment yet flexible enough to accommodate gradual modification.

Third, the framework provides a way to interpret processes of secularization and institutional change. When religious symbolic systems lose authority, the integrative representation that previously stabilised the BB-domain may weaken. Societies must then redistribute the symbolic functions that sustain coordination across other components of the BB-domain. Legal systems, bureaucratic governance, scientific authority, and secular moral norms may increasingly participate in maintaining shared reference structures. These institutions are not unique to secular societies, but their relative authority within the symbolic configuration can shift over time. Sociological analyses of modern societies have shown that religious authority can gradually be displaced by alternative symbolic institutions without eliminating the need for shared normative structures (Berger, 1967; Taylor, 2007). From the perspective developed here, secularization therefore represents a reconfiguration of the BB-domain rather than the disappearance of the underlying control architecture.

A further implication concerns the philosophical interpretation of the concept of God. Spinoza proposed that God and nature are identical, suggesting that the divine corresponds to the totality of reality itself (Spinoza, 2000). The framework developed here offers a structurally related but more restricted interpretation. Rather than identifying God with the entirety of existence, the present model interprets God as representing the totality of the BB-domain. In this view, God symbolizes the complete population-level configuration of shared symbolic representations that regulate individual cognition through the societal control loop.

This comparison suggests a reinterpretation of Spinoza's idea in terms of the dual architecture described earlier. Spinoza's formulation collapses all domains of reality into a single unified substance. By contrast, the BWBB framework distinguishes between the analog domain of brain–world interaction and the digital domain of brain–brain interaction. Within this dual architecture, the symbolic representation of God corresponds specifically to the BB-domain rather than to the entirety of the physical world. This reinterpretation preserves the intuition that God represents an integrated totality, while grounding that totality in the population-level symbolic system that organizes human coordination.

Finally, the framework generates empirical predictions. Societies in which symbolic systems strongly reinforce both overt and covert feedback channels should exhibit lower coordination error and greater long-term stability. Conversely, systems in which symbolic authority becomes fragmented or weakly

internalized should experience greater divergence among internal models and increasing difficulty sustaining large-scale coordination.

Taken together, these implications suggest that the persistence of complex human societies depends not only on material resources or institutional design but also on the stability of the symbolic structures that organise the BB-domain.

8. Conclusion

This paper proposes a control-theoretic account of how large human societies maintain coordination among cognitively autonomous individuals. The central problem is that individual cognition operates locally, while societal stability requires convergence among distributed expectations, norms, and predictions across large populations. Without mechanisms that align internal models across agents, coordination error accumulates and collective action becomes fragile.

The analysis addressed this problem by extending the BWBB framework. Within this architecture, cognition operates in two coupled domains. The BW-domain governs analog brain–world interactions through continuous sensorimotor feedback loops. The BB-domain governs digital brain–brain interactions through symbolically structured communication. While BW-level control stabilises individual behaviour relative to the physical environment, BB-level dynamics regulate alignment among individuals by stabilising shared symbolic representations.

Building on this architecture, the paper introduces a formal model of the BB-domain as a population-level dynamical system. In this model, individuals maintain internal models $M_i(t)$ that generate digitally structured contributions $I_i(t)$ to the shared symbolic state $S(t)$. The BB-domain integrates these distributed contributions and feeds back normative constraints onto individuals through overt and covert channels. Population-level coordination depends on maintaining low variance among internal models relative to the shared symbolic configuration.

Religious systems were analysed as historically persistent implementations of this control architecture. By stabilising symbolic structures, integrating overt sanctioning with covert internalized monitoring, and transmitting normative representations across generations, religious traditions can maintain convergence among distributed internal models in large populations.

Within this framework, the concept of God can be interpreted as the symbolic representation of the integrative wholeness of the BB-domain. The distributed regulatory structure that aligns individual cognition is represented as a unified evaluative authority that observes, judges, and guides behaviour. This representation compresses the complexity of distributed social expectations into a cognitively tractable form, strengthening both overt and covert feedback channels and reinforcing the stability of the societal control loop.

This interpretation does not depend on the metaphysical status of the entity represented. The stabilising function arises from the role the representation plays within individual internal models and within the population-level symbolic configuration. When individuals treat the symbolic authority as real and authoritative, the representation helps sustain the closed feedback loop between distributed agents and the shared normative state.

The framework developed here suggests that the persistence of complex human societies depends not only on material resources or institutional arrangements but also on the stability of the symbolic structures that organise the BB-domain. By modelling these structures as components of a population-level control system, the present account provides a unified way to analyse religion, institutions, and secular symbolic systems as mechanisms that regulate coordination among large groups of individuals.

The present framework is intended as a conceptual and formal model rather than a complete empirical theory of religion or social behaviour. The proposed variables and dynamics require further operationalization and empirical testing. Future work should examine how symbolic feedback can be quantified experimentally and how different forms of symbolic authority influence the alignment of internal models across individuals. Understanding how these symbolic regulatory structures emerge, stabilise, and transform remains an important task for future research on the dynamics of human societies.

Declarations

AI usage

Large language models (LLMs) were used to assist with literature discovery and to improve the clarity, organisation, and language of the manuscript. The scientific concepts, interpretations, and conclusions are those of the author.

Competing interests

The author declares no competing interests.

Ethics declaration

Ethical approval

was not required because this research did not involve human participants or their data.

Author Contribution

E.A. conceived the theoretical framework, developed the formal analysis, and wrote the manuscript.

Acknowledgements

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement No. 786949).

References

1. Ahissar E, Assa E (2016) Perception as a closed-loop convergence process. eLife 5:e12830
2. Ahissar E, Polani D, Ahissar M (2025) Digital-Analog Perceptual Duality. *Psychological review, Advance online publication* (https://dx.doi.org/10.1037/rev0000579)
3. Ashby WR (1956) An introduction to cybernetics. Chapman & Hall, London
4. Astrom KJ, Murray RM (2010) Feedback Systems: An Introduction for Scientists and Engineers. Princeton University Press
5. Atran S, Henrich J (2010) The evolution of religion: How cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. Biol Theory 5(1):18–30
6. Axelrod R (1984) The Evolution of Cooperation. Basic Books
7. Bellah RN (1967) Civil religion in America. Daedalus 96(1):1–21
8. Berger PL (1967) The Sacred Canopy: Elements of a Sociological Theory of Religion. Anchor Books, New York
9. Boyd R, Richerson PJ (2005) The Origin and Evolution of Cultures. Oxford University Press
10. Carlson JM, Doyle J (2002) Complexity and robustness. *Proceedings of the National Academy of Sciences, 99*(suppl_1), 2538–2545
11. Carver CS, Scheier MF (1998) On the Self-Regulation of Behaviour. Cambridge University Press
12. Cavalli-Sforza LL, Feldman MW (1981) Cultural transmission and evolution: A quantitative approach. Princeton University Press
13. Clark A (1998) Being there: Putting brain, body, and world together again. MIT Press
14. Clark A, Chalmers D (1998) The extended mind. *analysis, 58*(1), 7–19
15. Durkheim E (1995) The Elementary Forms of Religious Life. Free
16. Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. Am Econ Rev 90(4):980–994
17. Franklin GF, Powell JD, Emami-Naeini A (2015) Feedback Control of Dynamic Systems. Pearson
18. Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11(2):127–138
19. Hardin G (1968) The Tragedy of the Commons. Science 162(3859):1243–1248
20. Henrich J (2015) The Secret of Our Success. Princeton University Press

21. Henrich J, Ensminger J, McElreath R, Barr A, Barrett C, Bolyanatz A, Ziker J (2010) Markets, religion, community size, and the evolution of fairness and punishment. Science 327(5972):1480–1484
22. Hutchins E (1995) Cognition in the Wild. MIT Press
23. List C, Pettit P (2011) Group Agency: The Possibility, Design, and Status of Corporate Agents. Oxford University Press
24. Luhmann N (1995) Social Systems. Stanford University Press
25. Norenzayan A (2013) Big Gods: How Religion Transformed Cooperation and Conflict. Princeton University Press
26. North DC (1990) Institutions, Institutional Change and Economic Performance. Cambridge University Press
27. Olson M (1965) The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press
28. Ostrom E (1990) Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press
29. Parsons T (1951) The Social System. Free
30. Powers WT (1973) Feedback: beyond behaviorism. Science 179(71):351–356
31. Purzycki BG, Henrich J, Apicella CL, Atkinson QD, Baimel A, Cohen E, Norenzayan A (2016) Moralistic gods, supernatural punishment and the expansion of human sociality. Nature 530(7590):327–330
32. Searle JR (1995) The Construction of Social Reality. Free
33. Shariff AF, Norenzayan A (2011) Mean gods make good people: Different views of God predict cheating behaviour. Psychol Sci 22(6):803–809
34. Simon HA (1957) Models of Man. Wiley
35. Sosis R, Alcorta C (2003) Signalling, solidarity, and the sacred: The evolution of religious behaviour. Evol Anthropol 12(6):264–274
36. Spinoza Bd (2000) Ethics. Oxford University Press, Oxford
37. Taylor C (2007) A Secular Age. Harvard University Press, Cambridge, MA
38. Theiner G, Allen C, Goldstone RL (2010) Recognizing group cognition. Cogn Syst Res 11(4):378–395
39. Tomasello M (2007) Shared intentionality: The roots of human cooperation. Dev Sci 10(1):121–125
40. Tomasello M (2016) A Natural History of Human Morality. Harvard University Press
41. Turchin P (2016) Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth. Beresta Books
42. Tyler TR (2006) Why People Obey the Law. Princeton University Press
43. Von Foerster H (2003) Understanding understanding: Essays on cybernetics and cognition. Springer
44. Whitehouse H, François P, Savage PE, Currie TE, Feeney K, Cioni E, Turchin P (2019) Complex societies precede moralizing gods throughout world history. Nature 568(7751):226–229

45. Wiener N (1948) Cybernetics: Or Control and Communication in the Animal and the Machine. MIT Press
46. Young HP (1996) The economics of convention. J Economic Perspect 10(2):105–122
47. Young HP (1998) Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton University Press, Princeton, NJ

Figures

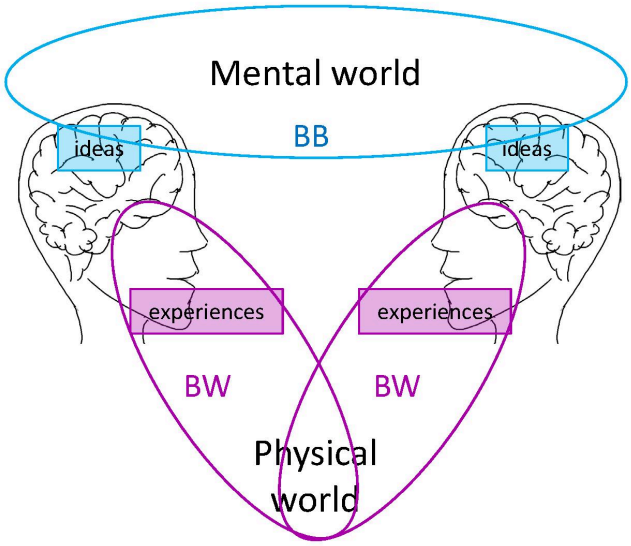


Fig 1. In human societies, individual subjects perceive and regulate their environments through two complementary domains: the brain–world (BW) domain, which links brains with the physical environment through sensorimotor experience, and the brain–brain (BB) domain, which links different brains through shared ideas, symbols, and concepts.

Figure 1

See image above for figure legend.

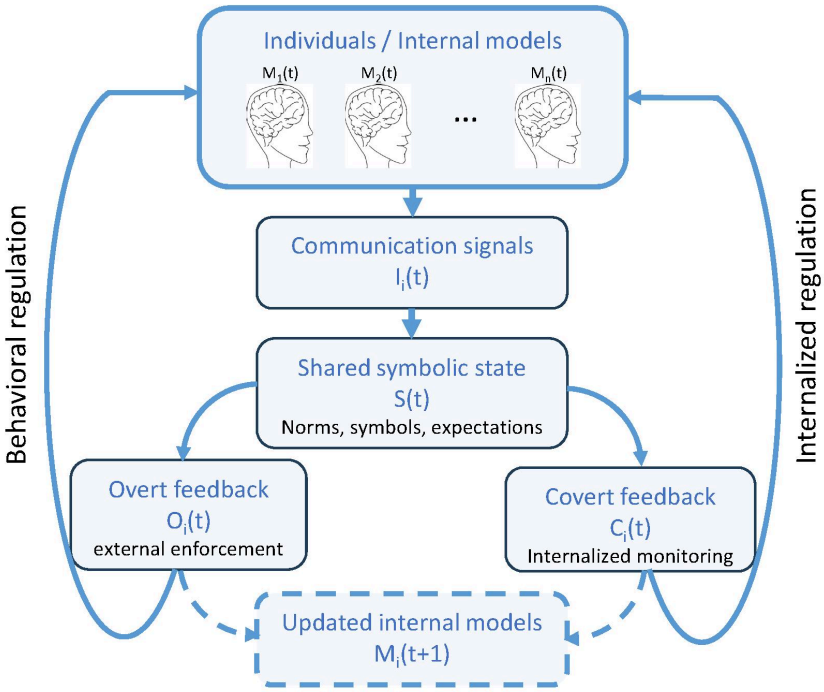


Fig 2. Closed symbolic feedback loop underlying social coordination. Individuals maintain internal models $M_i(t)$ that generate communicative signals $I_i(t)$, which collectively update a shared symbolic state $S(t)$ consisting of norms, symbols, and shared expectations. The shared symbolic state regulates behavior through overt feedback $O_i(t)$, including explicit rules and sanctions, and through covert feedback $C_i(t)$, including internalized monitoring and anticipatory evaluation. These feedback pathways produce updated internal models $M_i(t+1)$, marked here by dashed arrows and boxes to indicate temporal overlay in the graph.

Figure 2

See image above for figure legend.

Control term	Social implementation	Mechanism	timescale	examples
Proportional (P)	Overt feedback (external enforcement)	Detect deviation -> apply sanction	Short (immediate)	Laws, punishment, fines, social disapproval
Derivative (D)	Covert feedback (Internalized monitoring)	Anticipate deviation -> self-correct	Medium (near future)	Guilt, shame, conscience, anticipated reputation
Integral-like (I)	Accumulated social consequences	Long-term outcomes accumulate -> reshape beliefs and incentives	Long (extended)	Economic stability, prosperity, trust, resource availability

Fig. 3. Control-theoretic interpretation of symbolic social regulation. The principal regulatory mechanisms proposed in the present framework are mapped onto their analogous components in classical feedback control. **Proportional (P) control** corresponds to overt regulation through explicit rules, sanctions, and external enforcement that generate corrective responses to observed deviations from shared norms. **Derivative (D) control** corresponds to covert regulation through internalized monitoring, allowing individuals to anticipate potential deviations and initiate self-correction before overt violations occur. **Integral-like (I) regulation** corresponds to the accumulation of long-term social consequences, whereby sustained coordination, or its failure, gradually reshapes beliefs, incentives, trust, and resource availability. Together, these three components provide complementary mechanisms through which symbolic systems stabilize coordination in large human societies.

Figure 3

See image above for figure legend.