

The Mind as an Operating System: A Modular Governance Account of Internal Conflict and Agency

Classical accounts of agency presuppose psychological unity: a single subject or executive centre responsible for deliberation, normativity, and action. Internal conflict is then read as a deficiency—weakness of will, irrationality, failed self-control. This paper argues the assumption is mistaken. Drawing on cognitive science’s picture of cognition as distributed across semi-autonomous subsystems, on the clinical record of stable and structured internal conflict, and on the persistence of felt unity even under deep division, I propose that coherent agency does not require unity but governance.

On the modular governance model, the self is not a substance, a narrative centre, or a controlling module. It is a governance regime over a plurality of specialised subsystems—affect, valuation, vigilance, norm-enforcement, narrative integration, motivational projection—each with partially independent priorities and informational access. The familiar sense of a continuous “I” is the output of successful governance, not its precondition. I develop the proposal through the analogy of an operating system: as an OS does not perform applications’ tasks but governs their interaction through scheduling, access control, and conflict-resolution policies, the self does not originate motivations but regulates how subsystems compete and cooperate. Internal “myths”—pre-reflective normative policies—function as the system’s coherence and security rules: invisible in stable functioning, decisive under conflict.

The model’s distinctive claim concerns what makes governance authentic, and how it fails. Arbitration is the subject’s own only when the affective core retains access to the executive position; this access, not coherence as such, is the mark of authentic agency. It follows that the gravest failure of agency is not fragmentation but capture: an installed configuration—imprinted from caregivers and culture, or built defensively—can occupy the executive position and govern in the subject’s place with the affective core gated out, producing experience that is fully coherent and felt as authentic from within while no longer being the agent’s own. The paradigm is the high-functioning individual who harms systematically, narrates an irreproachable life, and registers no distress—precisely the profile that content-targeting therapies cannot reach, because they negotiate with the very structure that should be displaced.

This reframing has empirical bite. It distinguishes three governance regimes—Sovereign, Instable, Colonized—and predicts that neural and behavioural signatures sort by regime rather than by DSM category: the same diagnosis under different regimes should yield different signatures, and the same regime across diagnoses should yield one signature. And it relocates responsibility without dissolving it: explaining the architecture of harm does not excuse it.

The poster sets out the architecture, the regime taxonomy, and the falsifiable predictions, and shows why a governance framework reaches cases that both unity-based and symptom-based accounts systematically miss. The guiding question shifts from “what does this person feel?” to “what regime is this system running—and what would it take to restore sovereignty?”

Keywords: modular cognition; agency and self-control; internal conflict; governance models of the self; philosophy of mind; cognitive architecture

Selected references

Bratman, M. E. (2007). *Structures of Agency*. Oxford University Press.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.

Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Co.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*.

Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.