NUGAE - A PANDEMONIUM MODEL OF MORAL CHOICES

Arturo Tozzi
Former Center for Nonlinear Science, Department of Physics, University of North Texas, Denton, Texas, USA
Former Computationally Intelligent Systems and Signals, University of Manitoba, Winnipeg, Canada
ASL Napoli 1 Centro, Distretto 27, Naples, Italy
tozziarturo@libero.it

For years, I have published across diverse academic journals and disciplines, including mathematics, physics, biology, neuroscience, medicine, philosophy, literature. Now, having no further need to expand my scientific output or advance my academic standing, I have chosen to shift my approach. Instead of writing full-length articles for peer review, I now focus on recording and sharing original ideas, i.e., conceptual insights and hypotheses that I hope might inspire experimental work by researchers more capable than myself. I refer to these short pieces as nugae, a Latin word meaning "trifles", "nuts" or "playful thoughts". I invite you to use these ideas as you wish, in any way you find helpful. I ask only that you kindly cite my writings, which are accompanied by a DOI for proper referencing.

Empirical studies increasingly show that moral decision-making arises from distributed and partially independent neural systems whose activity patterns vary dynamically with context. Rather than assuming a central moral faculty or hierarchical control, we propose a Pandemonium model in which moral judgment results from the real-time competition and cooperation among multiple neural subnetworks. Each subnetwork operates as a computational demon, evaluating the same perceptual and contextual inputs under its own objective function. The outcome of this collective dynamic is not a predetermined rule, but rather an emergent equilibrium reflecting the transient dominance of specific evaluative dimensions.

Our multi-agent model accounts for observed neural co-activations and behavioral variability across contexts without postulating a central executive structure. Each demon corresponds to a neural system with a defined moral function and can be mathematically described by its cost function $C_k(s,a)$, where s is the perceived state and a the possible action. The overall moral output is determined by the weighted sum:

$$J(s,a) = \Sigma_k \lambda_k(s) \cdot C_k(s,a)$$

where $\lambda_k(s)$ are context-dependent weights subject to the normalization constraint $\Sigma_k \lambda_k(s) = 1$. Decision probabilities follow a Boltzmann distribution:

$$P(a|s) = \exp[-J(s,a)/T] / \Sigma_{a'} \exp[-J(s,a')/T]$$

where T models stochastic noise or emotional arousal. Gating weights evolve through a softmax rule:

$$\lambda_k(s) = \exp[g_k(s)/\tau] / \Sigma_j \exp[g_j(s)/\tau]$$

in which $g_k(s)$ encodes neural activation strength and τ defines competition intensity.

Below are five demons with their corresponding brain regions and cost functions:

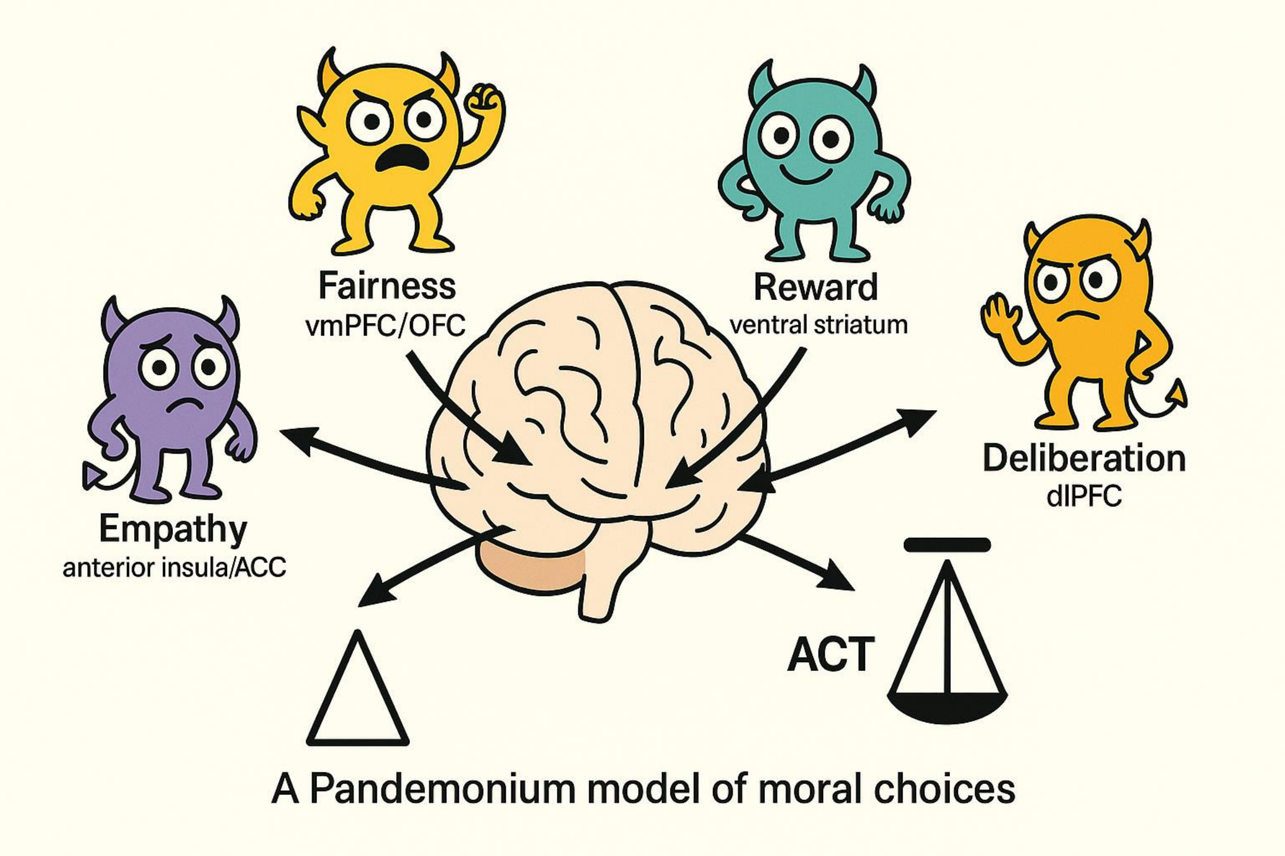
1. Empathy demon (anterior insula, anterior cingulate cortex):
$$C_{emp}(s,a) = \Sigma_i w_i \cdot h_i(s,a)$$
where h_i represents harm prediction for agent i and w_i scales interpersonal proximity.
2. Fairness demon (ventromedial and orbitofrontal prefrontal cortex):
$$C_{fair}(s,a) = (1/N)\, \Sigma_i [u_i(s,a) - \bar{u}(s,a)]^2$$
where u_i is expected utility.
3. Reward demon (ventral striatum, nucleus accumbens):
$$C_{rew}(s,a) = -[u_{self}(s,a) + \alpha \cdot A(s,a)]$$
where A(s,a) is predicted approval.
4. Authority demon (amygdala, hippocampus, temporoparietal junction):
$C_{auth}(s,a) = KL(\delta_a \parallel \pi_R(\cdot\, |s))$
quantifying deviation from the normative distribution π_R.
5. Deliberation demon (dorsolateral and frontopolar prefrontal cortex):
$$C_{delib}(s,a) = -V_\gamma(s,a) + \eta \cdot E(s,a)$$
where E(s,a) denotes cognitive cost.

Each demon thus contributes a partial evaluation, and the final moral decision emerges from their weighted competition without postulating a central executive structure.

The Pandemonium framework can be experimentally validated by correlating fMRI or electrophysiological signatures of each demonâ€™s anatomical substrate with predicted weight values λ_k. Dynamic causal modeling could estimate coupling between subnetworks during moral dilemmas, while lesion-inspired manipulations in silico could predict altered moral outcomes following selective suppression (e.g., $\lambda_{emp} = 0$). Quantitative behavioral predictions include condition-specific shifts in act/omit probabilities that correspond to empirically measurable changes in activation within empathy and authority circuits

Our Pandemonium model transforms moral cognition into a computable, testable system linking mathematical structure and neural activity. Its implications extend across domains.
In neuroscience, it provides a formal mapping between cost functions and cortical–subcortical dynamics, guiding hypothesis-driven neuroimaging studies.
In psychology, it enables estimation of individual moral profiles through fitted λ_k values, revealing stable moral signatures analogous to cognitive styles.
In psychiatry, moral deficits could be reframed as distortions in weight distributions: excessive authority weighting in obsessive–compulsive disorder, reduced empathy weighting in psychopathy, or unstable gating in frontal syndromes. In artificial intelligence, our model provides a moral decision structure capable of context-sensitive negotiation among competing goals rather than rigid rule following. It could inform architectures for socially aligned autonomous systems, integrating normative and affective reasoning.
In philosophy and ethics, it operationalizes pluralistic moral theories, allowing computational exploration of how deontological and consequentialist components interact in real decisions.

Overall, the Pandemonium model converts moral plurality into a rigorous, measurable system that unites neuroscience, computation and ethics within a single theoretical structure, enabling future empirical testing and interdisciplinary development.


Fairness
vmPFC/OFC
Reward
ventral striatum
Empathy
anterior insula/ACC
Deliberation
dlPFC
ACT
A Pandemonium model of moral choices