

Large-Language Models as a Cognitive Virus

Ricard Solé,^{1,2,3,4,*} Giulio Ruffini,^{5,†} Francesca Castaldo,^{5,‡} Marco Tuccio,^{1,3} Luis F. Seoane,^{3,6}
Manlio de Domenico,^{7,8,9,10} Santiago F. Elena,^{11,4} David C. Krakauer,⁴ and Michael Levin^{12,13}

¹Complex Systems Lab, Universitat Pompeu Fabra (MELIS), Dr. Aiguader 88, 08003 Barcelona, Spain

²Institució Catalana de la Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

³Institut de Biologia Evolutiva, CSIC-UPF, Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain.

⁴Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM, United States.

⁵Barcelona Computational Foundation (BCOM) and Neuroelectronics, Barcelona, Spain

⁶Okinawa Institute of Science and Technology Graduate University, Onna, 904-0495, Japan.

⁷Complex Multilayer Networks Lab, Department of Physics and Astronomy

'Galileo Galilei', University of Padua, Via Marzolo 8, 35131 Padova, Italy

⁸Istituto Nazionale di Fisica Nucleare, Sez. Padova, 35131 Padova, Italy

⁹Padua Center for Network Medicine, Via Marzolo 8, 35131 Padova, Italy

¹⁰Padua Neuroscience Center, Via Giuseppe Orus, 2, Italy

¹¹Instituto de Biología Integrativa de Sistemas (I2SysBio), CSIC-Universitat de València, Paterna, València, 46980, Spain

¹²Allen Discovery Center, Tufts University, Medford MA, United States.

¹³Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston MA, United States.

Large-language models (LLMs) are rapidly becoming part of human culture, reshaping how information is produced, transmitted, and used. Here we propose that their diffusion can be understood through a viral analogy, with LLM use spreading through populations, becoming embedded in cognitive and cultural practices. We model transitions among uncoupled, coupled, and persistently dependent users, and show that the interplay between social transmission, recovery, and collective reinforcement can generate tipping points and technological lock-in. A central consequence is the possibility of runaway dynamics: once a critical threshold is crossed, small increases in adoption can trigger rapid population-level shifts toward persistent dependence, with abrupt losses in cognitive competence. The same framework, however, identifies conditions for cognitive immunization, based on reducing transmission and facilitating reversibility. Our results highlight how LLM adoption may involve nonlinear collective transitions with important consequences for cognitive autonomy.

Keywords: large-language models, language, cognitive offloading, extended mind, memes, automation bias, dependency, cognition

Significance Statement. Large language models are rapidly becoming embedded in everyday cognitive and cultural practices, yet we still lack simple theoretical frameworks for understanding how their adoption may reshape collective patterns of cognition. Here we introduce the idea of LLMs as cognitive viruses: culturally transmitted technologies whose spread is promoted by their usefulness but can also increase dependence through cognitive offloading. A dynamical model shows that social facilitation can generate runaway transitions from predominantly autonomous cognition to widespread technological dependence. These results suggest that abrupt collective changes can emerge even when individual adoption is gradual. By connecting epidemic dynamics, cultural transmission, and cognitive offloading, our framework identifies general mechanisms that may govern human-AI coevolution and provides a basis for “cognitive immunization”.

I. INTRODUCTION

The emergence of human language is widely regarded as a major evolutionary transition [1]. While it shares key features with genetic transmission [2], language established a new system of inheritance and collective memory. It allows information to accumulate, be passed on, and recombine over generations, thus supporting cumulative culture [3, 4]. In this sense, languages can be viewed as population-level cultural systems with partially autonomous dynamics: they diversify, compete, hybridize, spread, and sometimes become extinct, showing patterns reminiscent of species embedded in ecological communities [5]. Their persistence depends largely on social learning, through which individuals acquire words, grammatical constructions, meanings, and communicative conventions by observing and interacting with others, especially during childhood [6]. Because linguistic structures propagate by entering developing minds, relying on their exceptional plasticity, and recruiting their speakers as new vectors of transmission, language has sometimes been described as a viral entity [7, 8].

The metaphor acquired a concrete technological counterpart with the emergence of computer viruses: self-propagating informational structures capable of enter-

* Corresponding author, ricard.sole@upf.edu

† Corresponding author, giulio.ruffini@bcom.one

‡ Corresponding author, francesca.castaldo@bcom.one

ing a host system, modifying its operation, and recruiting it for further transmission [9]. In parallel, theories of cultural evolution formalized the idea that socially transmitted information can exhibit variation, differential persistence, and inheritance, whether framed through gene-culture coevolution [10–12] or through the concept of memes [13–15]. Together, these traditions point towards a broader class of entities, the “viruses of the mind” [16, 17] whose evolution depends not on a particular material substrate, but on their capacity to reproduce and persist across biological, cognitive, or computational hosts.

In language, what propagates virally are patterns—from words to ideas—that reproduce through imitation, mutate through use, and persist through brains, media, institutions, and technologies. This points to cognition as distributed across minds and external structures—the extended mind [18]—which do not merely store information but also reshape cognitive tasks [19, 20]. Such artifacts can be *complementary*, strengthening capacities beyond their immediate use, or *competitive*, improving performance while potentially weakening the underlying skill [21–23]. Language itself is a two-sided cognitive technology: it externalizes, stabilizes, and recombines thought [24], while linguistic categories and narratives reshape perception and reasoning [25–28]. Through cultural transmission, language is also adapted to the cognitive and communicative demands of its users [29–31], so that its propagation both extends cognition and transforms the minds that depend on it.

This reciprocal transformation also characterizes communication technologies. The Internet and instant messaging blurred distinctions between speech and writing [32, 33], while mobile phones and smartphones encouraged abbreviated, compressed forms of expression that largely reflect linguistic flexibility rather than declining literacy [34–36]. Social media further reorganized the structure and social selection of communication [37, 38], integrating text, hyperlinks, multimodal expression, and algorithmically selected content [39, 40]. Such technologies do not merely transmit language: their affordances reshape how it is produced, circulated, interpreted, and selected.

LLMs represent a further transition. Unlike previous media, they do not merely constrain the form or circulation of human expression but actively participate in its production, reformulation, and evaluation. The relevant question, therefore, is not simply whether LLMs alter language use, but whether they reorganize the coupled system formed by language, cognition, and technology. This comparison may help identify which cognitive capacities these systems amplify, which they displace, and under what conditions sustained reliance may produce cognitive reorganization or erosion. In this respect, LLMs extend a feature shared, to varying degrees, by earlier information technologies: cognitive offloading, understood as the delegation of cognitive operations to external tools and representations.

Two technological precursors bridge natural language and the LLM era: computer viruses (CVs) [41–43] and programming languages (PLs) [44]. CVs reproduced key features of biological viruses: compact informational structures that exploit host machinery and evolve to evade detection. Their emergence triggered an arms race with antivirus software [45], influenced artificial life [46] and network science [47–49], and fostered modern cybersecurity and computer immunology [50, 51]. PLs spread across computers, institutions, and communities [52, 53], becoming an operational substrate of the digital revolution and, alongside neural networks, paving the way for artificial intelligence [54]. Both depend on technological hosts and reshape their environments. LLMs combine aspects of each: like PLs, they provide a new human-machine interface; like CVs, they propagate through users and digital ecosystems. Yet their capacity for cognitive offloading is unprecedented [55].

Here, we address our previous question by treating LLMs as engines of cognitive and social change whose rapid diffusion is transforming human capacities. We first compare LLMs with infectious agents and then develop an explicit population model of their propagation, drawing on mathematical approaches from epidemiology [56] that have also been applied to technological adoption [57, 58]. The model describes transitions among host-coupling states, which we link to an illustrative measure of cognitive competence to distinguish the consequences of different coupling regimes from the underlying bifurcation structure and explore potential interventions. It does not identify any single entity as the viral analogue: LLM ecosystems contain culturally transmitted practices and content alongside technologically evolving model lineages, and these need not coincide. Instead, the analogy concerns a broader feedback loop in which persistent technological lineages are instantiated in external machinery, modify their human host environment, and thereby influence their own propagation. This homology is already partly present in conventional and adaptive software [59–62], but LLMs deepen it by becoming integrated into cognitive processing itself.

The viral analogy does not imply that LLM-human interactions are intrinsically parasitic. Biological viruses range from pathogens to mutualists and evolutionary partners [63–67], and the effects of LLMs likewise depend on how they are used. They can enhance exploration, access to expertise, and productivity [68, 69], but can also promote cognitive offloading, dependence, and loss of competence [70], while individual benefits may not scale to the population level [71]. The analogy is therefore useful because LLM use can generate persistence and transmission loops, both through culturally propagating patterns of use and through model lineages sustained and modified by technological and economic feedbacks. The model below focuses on one part of this ecology: transitions among human states of cognitive coupling.

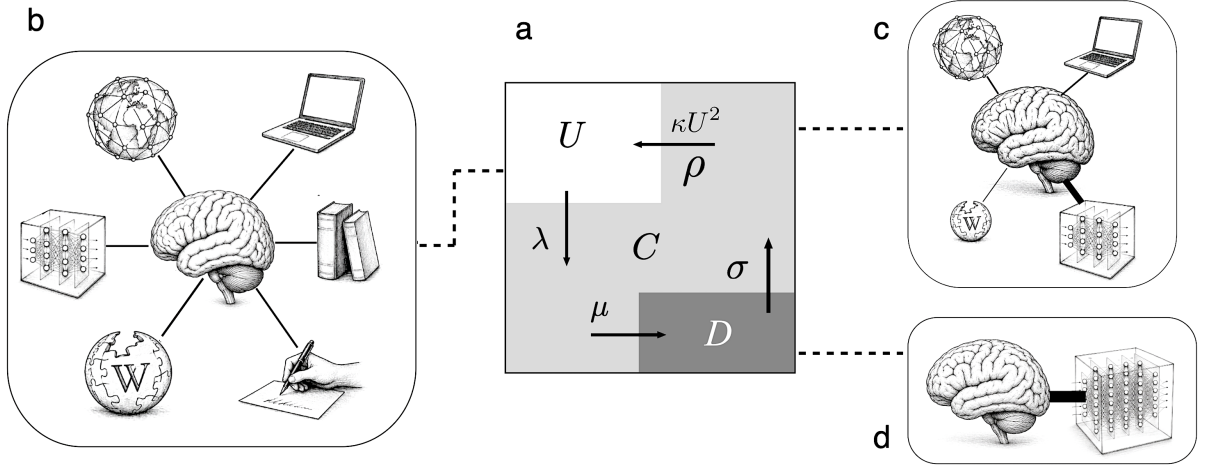


FIG. 1. Population-level host-state model within an LLM propagation ecology. (a) Individuals move among three states: uncoupled or weakly coupled users U ; autonomous coupled users C , who use LLMs while retaining reading, writing, reasoning, verification, and access to alternative information sources; and persistently dependent users D , for whom LLM-mediated cognitive operations have become strongly substitutive. Exposure to socially, institutionally, or platform-mediated LLM practices drives the transition $U \rightarrow C$ at rate λ , whereas abandonment returns users to the uncoupled state at rate ρ . Regular use develops into dependency at rate μ , while training, verification practices, or deliberate cognitive friction restore autonomous use at rate σ . (b) In the uncoupled state, cognition is distributed across the brain and a heterogeneous external information environment, including books, writing, computers, Wikipedia, and the Internet. (c) In the coupled but autonomous state, the LLM becomes an additional component of this cognitive ecology, while other information sources and independent cognitive operations are retained. (d) In the dependent state, interaction with the LLM becomes dominant, progressively replacing alternative cognitive supports and concentrating cognitive activity within a single human-machine coupling. The model therefore describes transitions among human host/coupling states associated with LLM use. It represents one population-level layer of the broader viral ecology and does not by itself model reproduction of the model/program lineage.

II. POPULATION DYNAMICS OF LLM-MEDIATED COGNITIVE COUPLING

Classical epidemic models describe the transmission of biological agents through transitions between host states [56]. A similar mathematical language has been used for technological diffusion, where adoption depends on exposure to previous users and social reinforcement [57, 72], as well as in multicompartment models of drug [73–75] and social media [76–78] addictions. At the level modeled here, we do not represent the reproduction or evolution of the model/program lineage explicitly. Instead, we coarse-grain one epidemiological layer of the larger LLM ecology: transitions among human host/coupling states. We consider three population states (Fig. 1a): uncoupled or weakly coupled individuals U , regular users C who retain cognitive autonomy, and dependent users D who persistently delegate cognitive operations to the model. Their dynamics are

$$\frac{dU}{dt} = -\lambda UC + \rho C + \kappa U^2 C, \quad (1)$$

$$\frac{dC}{dt} = \lambda UC - (\mu + \rho)C + \sigma D - \kappa U^2 C, \quad (2)$$

$$\frac{dD}{dt} = \mu C - \sigma D, \quad (3)$$

with $U + C + D = 1$. Here, λ measures the effective spread of LLM practices through social and institutional exposure, ρ the return from regular use to the

uncoupled state, μ the transition from regular use to persistent dependency, and σ recovery from dependency. These rates coarse-grain heterogeneous individual processes into population-level transitions rather than representing single psychological mechanisms. The nonlinear term $\kappa U^2 C$ introduces a cooperative (Allee-like) mechanism [79]. The underlying assumption is that LLM-independent cognitive practice is socially reinforced and cultural expectations that reward independent reasoning become more effective when autonomous individuals are common. The quadratic dependence on U represents this positive frequency dependence, whereas the factor C restricts the restoring effect to individuals engaged with LLMs. Accordingly, the incidence term λUC should be interpreted as an effective host-side social or institutional transmission pressure. As in biological compartment models, the fact that the state variables describe hosts does not imply that a host state is the pathogen. The equations model the epidemiology of coupling and therefore do not, by themselves, determine the identity of the viral analogue.

Mean-field approximations like the one we take here provide a valuable, analytically tractable baseline for identifying the mechanisms underlying collective behavior [80]. Here, individuals sample population-level frequencies, with connectivity absorbed into the effective rate λ . However, real social and technological networks are heterogeneous and correlated [81]: connectivity distributions can shift epidemic thresholds and immuniza-

tion outcomes [47, 82–86], while other network properties can further modify diffusion [87–91]. Communication topology could also shape collective behavior in LLM-agent networks (as they do in biological epidemics [92]), particularly when cognitive states reshape social ties (again, as for biological pathogens [93]). At larger scales, interconnected regions and institutions can generate invasion thresholds and topology-dependent patterns [94–96]. Our formulation therefore offers a useful reference for future network and spatial extensions.

The cooperative component at the population-level has strong precedents in social and ecological dynamics. The quadratic term, commonly used in evolutionary ecology models to introduce mutualisms [80, 97, 98] assumes that autonomous cognitive practices are frequency dependent: schools, workplaces, peer groups, and norms can make independent work easier to maintain when it remains common. Threshold models of collective behavior show theoretically that such frequency dependence naturally produces critical transitions in humans [99–101]. In particular, it has been shown that social conventions can undergo genuine tipping [102, 103] when a minority of the group was sufficient to move a population from one convention to another.

Dependency is assumed to arise primarily through regular use, $C \rightarrow D$, rather than directly from U . These simplifications isolate the interaction between contagion-like technological adoption and collective protection of cognitive autonomy, allowing us to identify conditions for abrupt and potentially irreversible population-level transitions. Figure 1 connects these population states with different forms of human-LLM coupling. Uncoupled individuals remain embedded in a diversified cognitive ecology, while making little or no use of LLMs (Fig. 1b). Regular users incorporate LLMs while retaining substantial independent cognitive activity (Fig. 1c), whereas in the dependent state the LLM becomes the dominant interface for performing tasks and accessing information (Fig. 1d).

Transitions among these regimes can result from social exposure, institutional adoption, repeated reliance on generated outputs, abandonment of the technology, or progressive cognitive offloading. Cognitive offloading is itself a normal and often adaptive component of human cognition [55, 70, 104]. Reading and writing provide a paradigmatic example: literacy recruits and reorganizes preexisting neural circuits [105], while external symbolic structures extend memory and enable new forms of reasoning. Crucially, however, external support can either generate new internal competence or substitutes for it.

Recent work on LLM-assisted writing suggests that increased externally supported performance can be accompanied by reduced cognitive engagement, recall, or sense of authorship [106]. It is therefore useful to distinguish between *scaffolding* and *substitution*. Scaffolding reduces immediate cognitive demands while preserving or increasing the user’s subsequent capacity to perform the task; substitution removes the need to perform the

underlying cognitive operation. Following [21], these correspond broadly to *complementary* and *competitive cognitive artifacts*. An LLM may thus operate as a tutor whose arguments are reconstructed, challenged, and verified, or as a substitute that performs synthesis, evaluation, and composition on behalf of the user. The relevant distinction is therefore not simply between using and avoiding LLMs, but between forms of coupling and, ultimately, what cognitive capacities remain when the tool is withdrawn. The transition $C \rightarrow D$ provides a minimal representation of this shift from complementary use toward persistent substitutive dependency.

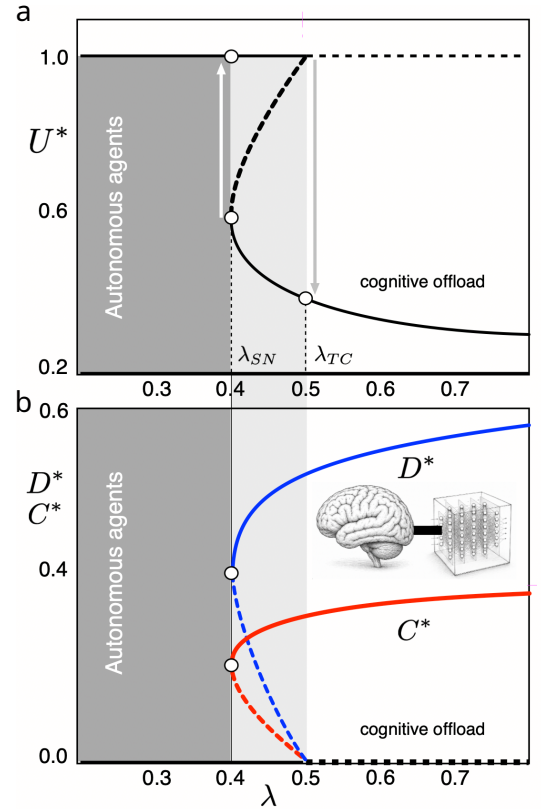


FIG. 2. Bifurcation structure of the LLM propagation model. (a) Equilibrium fraction of uncoupled individuals, U^* , as a function of the transmission parameter λ . For $\lambda_{SN} < \lambda < \lambda_{TC}$, the system is bistable: the fully uncoupled state and a coupled state are both stable and are separated by an unstable branch (dashed). Increasing λ drives an abrupt transition to the coupled regime at the transcritical threshold λ_{TC} , whereas recovery occurs only after λ is reduced below the saddle-node threshold λ_{SN} , generating hysteresis and technological lock-in. (b) Corresponding equilibrium fractions of regular users, C^* (red), and dependent users, D^* (blue). Solid curves denote stable equilibria and dashed curves unstable branches. Beyond the tipping point, the sharp decrease in U^* is accompanied by a rapid increase in both regular LLM use and persistent cognitive dependency, representing population-level cognitive offloading. Parameters used: $\rho = 0.10, \kappa = 0.40, \mu = 0.20, \sigma = 0.10$.

The model displays qualitatively distinct long-term regimes of LLM use. Using $D = 1 - U - C$, the dynamics

can be written as a two-dimensional system

$$\frac{dU}{dt} = Cf(U), \quad (4)$$

$$\frac{dC}{dt} = -Cf(U) - \mu C + \sigma(1 - U - C), \quad (5)$$

where $f(U) = \rho - \lambda U + \kappa U^2$. An equilibrium (fixed point) is the fully uncoupled state $E_U = (1, 0, 0)$, corresponding to a population in which LLM-mediated cognitive coupling is absent. The other points satisfy $f(U^*) = 0$ and $D^* = (\mu/\sigma)C^*$, giving

$$C_{\pm}^* = \frac{\sigma}{\mu + \sigma}(1 - U_{\pm}^*), \quad D_{\pm}^* = \frac{\mu}{\mu + \sigma}(1 - U_{\pm}^*), \quad (6)$$

where $U_{\pm}^*$ are the two roots defined as:

$$U_{\pm}^* = \frac{\lambda \pm \sqrt{\lambda^2 - 4\kappa\rho}}{2\kappa}. \quad (7)$$

The two branches exist when $\lambda \geq \lambda_{\text{SN}} = 2\sqrt{\kappa\rho}$. The uncoupled equilibrium remains stable while

$$\lambda < \lambda_{\text{TC}} = \rho + \kappa \quad (8)$$

and loses stability at λ_{TC} through a transcritical bifurcation. When $\kappa > \rho$, the saddle-node occurs before this loss of stability, producing the bistable interval

$$2\sqrt{\kappa\rho} < \lambda < \rho + \kappa. \quad (9)$$

Within this interval, both the uncoupled and the coupled state coexist, separated by the unstable branch (dashed curves in Fig. 2). Consequently, increasing λ from the uncoupled state leaves the population near $U = 1$ until λ_{TC} is reached, whereas decreasing λ from the coupled state does not restore the uncoupled regime until λ_{SN} is crossed. The difference between these thresholds,

$$\lambda_{\text{TC}} - \lambda_{\text{SN}} = (\sqrt{\kappa} - \sqrt{\rho})^2, \quad (10)$$

defines the width of the hysteretic region and provides a simple mechanism for technological lock-in.

These transitions are illustrated in Fig. 2. Figure 2a shows the equilibrium uncoupled fraction U^* as λ is varied: a stable coupled branch and a saddle emerge at λ_{SN} , while the uncoupled state loses stability only at λ_{TC} . The corresponding regular-use and dependent fractions are shown in Fig. 2b and follow directly from Eq. (6). Figure 3b summarizes the bifurcation structure in the (λ, κ) plane. Genuine bistability occurs only for $\kappa > \rho$ and between the saddle-node boundary $\lambda = 2\sqrt{\kappa\rho}$ and the transcritical boundary $\lambda = \rho + \kappa$. For $\kappa < \rho$, the formal saddle-node lies outside the physical simplex and the transition is instead continuous, occurring through a forward transcritical bifurcation. At $\kappa = \rho$, the two thresholds coincide at $\lambda = 2\rho$, marking the boundary between continuous adoption and history-dependent discontinuous transitions.

III. COGNITIVE COMPETENCE ACROSS THE TRANSITION

The population-level transition can have consequences for human competence, but their sign is not fixed by the bifurcation itself. To illustrate one possible consequence, we assign each state a relative cognitive competence, Γ_u , Γ_c , and Γ_d , and define the average

$$\langle \Gamma \rangle = \Gamma_u U + \Gamma_c C + \Gamma_d D. \quad (11)$$

Here Γ should be narrowly interpreted as cognitive competence (CC): the capacity available to the human when external support is removed, rather than the total capability of the coupled human-AI system. To illustrate a substitutive-use regime, we use $\Gamma_u = 1$, $\Gamma_c = 0.5$, and $\Gamma_d = 0.1$, corresponding to progressively greater loss of cognitive competence with increasing dependence. This ordering is an illustrative modelling assumption, not a general claim about LLM use. In scaffolded or augmentative regimes, coupling could leave subsequent CC unchanged or even increase it.

Using the previous attractor states, the average cognitive competence becomes (see SM for a full derivation)

$$\langle \Gamma \rangle^* = g + (1 - g)U^*, \quad (12)$$

where

$$g = \frac{\Gamma_c \sigma + \Gamma_d \mu}{\mu + \sigma} \quad (13)$$

is the mean effective CC of the coupled population. Thus, $\langle \Gamma \rangle^*$ is an increasing function of U^* and inherits exactly the same saddle-node, bistable, and hysteretic structure as the population dynamics (Fig. 3a-b). In particular, when $\kappa > \rho$, increasing λ leaves the system on the fully uncoupled state, $\langle \Gamma \rangle^* = 1$, until $\lambda_{\text{TC}} = \rho + \kappa$. At this point the autonomous attractor loses stability and the population moves to the coupled branch, for which $U_{-}^* = \rho/\kappa$ at the transition. The corresponding drop in CC is therefore

$$\Delta\Gamma_{\text{loss}} = (1 - g) \left(1 - \frac{\rho}{\kappa}\right). \quad (14)$$

Conversely, when λ is decreased from the coupled regime, recovery does not occur at the same point. The system remains on the low-CC branch until the saddle-node $\lambda_{\text{SN}} = 2\sqrt{\kappa\rho}$ is reached, where $U^* = \sqrt{\rho/\kappa}$. The subsequent return to the autonomous state produces a CC increase

$$\Delta\Gamma_{\text{rec}} = (1 - g) \left(1 - \sqrt{\frac{\rho}{\kappa}}\right). \quad (15)$$

cognitive competence therefore displays the same hysteresis as LLM adoption: gradual changes in transmission pressure can generate an abrupt loss of autonomous capacity, while reversing that loss requires a larger reduction in λ .

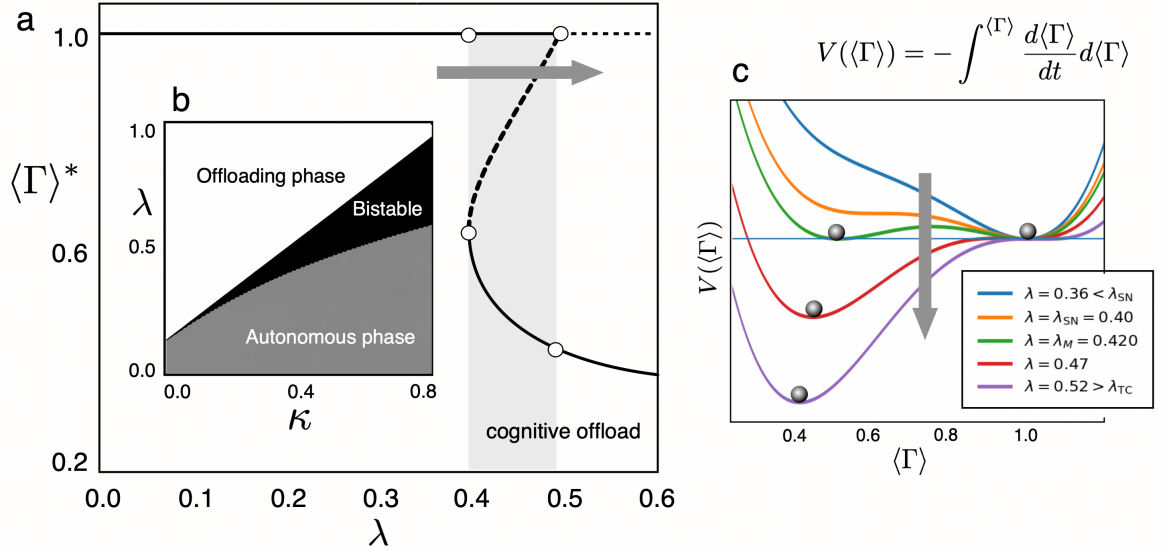


FIG. 3. Cognitive offloading bifurcation, phase structure, and effective potential of the LLM propagation model. (a) Equilibrium average cognitive competence $\langle \Gamma \rangle^*$ as a function of the transmission parameter λ . The fully autonomous state remains stable up to the transcritical threshold $\lambda_{TC} = \rho + \kappa = 0.50$, whereas a stable offloading branch and an unstable branch appear at the saddle-node threshold $\lambda_{SN} = 2\sqrt{\kappa\rho} = 0.40$. The shaded interval $\lambda_{SN} < \lambda < \lambda_{TC}$ marks the bistable region, where autonomous and offloaded states coexist, implying hysteresis and path dependence. (b) Phase diagram in the (κ, λ) plane. The lower boundary, $\lambda_{SN} = 2\sqrt{\kappa\rho}$, and the upper boundary, $\lambda_{TC} = \rho + \kappa$, delimit the bistable domain (black); below it the autonomous phase dominates (gray), whereas above it only the offloading phase is stable (white). (c) Effective potential $V(\langle \Gamma \rangle)$ for representative values of λ , following the gray arrow in (a) that crosses the bistable regime and finally by a single offloading minimum above λ_{TC} . At $\lambda = \lambda_M$ (the so called Maxwell point), the two minima have equal depth. Same parameters as in Fig. 2, with $\Gamma_u = 1$, $\Gamma_c = 0.5$, and $\Gamma_d = 0.1$.

The bifurcation structure itself is independent of this particular choice of Γ ; the values used here illustrate the consequences of a regime in which increasing dependence is associated with reduced CC. For the parameters used in Fig. 2, we obtain $g = 7/30 \simeq 0.233$. As λ increases through $\lambda_{TC} = 0.50$ (gray arrows), the equilibrium CC consequently falls from 1 to approximately 0.425. In the reverse trajectory, the coupled state persists to $\lambda_{SN} = 0.40$, where $\langle \Gamma \rangle^* \simeq 0.617$, before recovering to unity. The bifurcation therefore converts a smooth change in the pressure to adopt LLM-mediated cognition into a discontinuous and history-dependent change in population-level cognitive competence. Notice also that μ and σ do not alter the positions of the two tipping points in this minimal model, but they determine g and hence the cognitive cost associated with occupying the coupled state. Importantly, the presence of this sharp transition has a key implication: crossing the bifurcation changes the attractor structure and can therefore produce an abrupt change in the equilibrium population state under quasistatic parameter variation. The minimal model does not, however, determine how rapidly the transition unfolds in real time.

An alternative visualization of this transition is provided by the effective potential associated with the dynamics of $\langle \Gamma \rangle$. Under the one-dimensional reduction described in the SM, the coupled fractions C and D rapidly relax to their quasi-equilibrium ratio. The resulting dy-

namics can be written as a gradient system,

$$\frac{d\langle \Gamma \rangle}{dt} = - \frac{dV}{d\langle \Gamma \rangle}, \quad (16)$$

where the potential V is obtained from:

$$V(\langle \Gamma \rangle) = - \int^{\langle \Gamma \rangle} \left(\frac{d\Gamma}{dt} \right) d\Gamma \quad (17)$$

is a quartic function with two wells (see SM). This representation preserves the equilibrium states and bifurcation thresholds of the population model: minima of V correspond to stable population states, whereas the intervening maximum represents the unstable branch separating their basins of attraction.

The evolution of this landscape as λ increases is shown in Fig. 3c. The dynamics can be pictured as a marble moving over the landscape until it settles in a valley. For low values of λ , the only valley lies at high $\langle \Gamma \rangle$, corresponding to strong cognitive autonomy. At λ_{SN} , a second valley appears at lower $\langle \Gamma \rangle$, representing a state dominated by cognitive offloading. Within the bistable interval, $\lambda_{SN} < \lambda < \lambda_{TC}$, either state can persist, depending on the system's history and on which side of the intervening barrier it lies. As λ increases, the low- $\langle \Gamma \rangle$ valley becomes progressively deeper. At $\lambda_M \simeq 0.420$ (the so called Maxwell point) for the parameters used here, the two valleys have equal depth; beyond it, the offloading state becomes increasingly favored. Nevertheless, the

marble may remain trapped in the high-autonomy valley until this valley disappears at λ_{TC} . It then rolls abruptly toward the low- $\langle\Gamma\rangle$ minimum, producing a runaway transition: greater reliance on LLMs promotes further cognitive offloading, which in turn reinforces that reliance. In contrast, recovery of autonomy requires crossing the barrier or reducing λ sufficiently so that the offloading valley disappears at λ_{SN} . The landscape thus provides an intuitive picture of tipping, runaway change, and hysteresis between autonomous and cognitively offloaded states.

IV. COGNITIVE IMMUNIZATION AND INTERVENTION STRATEGIES

The bifurcation structure provides a natural framework for considering “cognitive immunization,” understood here not as preventing contact with LLMs, but as preserving resistance to harmful substitutive or dependency-producing coupling while allowing beneficial forms of human-AI integration. A first important result is the asymmetry between prevention and reversal. If the population is initially close to the uncoupled state E_U , invasion is prevented provided

$$\lambda < \lambda_{TC} = \rho + \kappa. \quad (18)$$

Once the system has moved to the coupled attractor, however, restoring λ below λ_{TC} is insufficient whenever the system lies inside the bistable regime. The coupled state persists until

$$\lambda < \lambda_{SN} = 2\sqrt{\kappa\rho}. \quad (19)$$

Thus, after the tipping point has been crossed, reversal requires a larger reduction in transmission pressure than would have been required to prevent the transition in the first place. This prevention-reversal asymmetry is the direct dynamical consequence of hysteresis and provides a simple mechanism for technological lock-in.

Among the parameters controlling the bifurcation structure, ρ has a particularly clear protective role. Increasing ρ , the rate at which regular users return to uncoupled or weakly coupled cognition, raises the invasion threshold λ_{TC} while simultaneously moving the system toward the boundary $\rho = \kappa$. At this boundary the physically accessible saddle-node disappears; for $\rho \geq \kappa$, bistability is lost and the transition becomes continuous. Interventions that make autonomous cognition an accessible and recurrent state—for example, through protected unaided tasks, deliberate periods of disengagement, maintenance of non-LLM skills, or attractive non-LLM alternatives—can therefore modify not only the position of the tipping point but the qualitative form of the transition itself.

The effect of κ is more subtle. Increasing κ strengthens the collective reinforcement of autonomous cognition and raises λ_{TC} , making invasion more difficult when the population is predominantly uncoupled. At the same time,

for $\kappa > \rho$, it also increases the separation between the two thresholds, $\Delta\lambda = \lambda_{TC} - \lambda_{SN}$ (Eq. 10), thereby enlarging the hysteretic region. Strong collective protection can therefore stabilize autonomy while U is high, but may also increase path dependence once the autonomous fraction has been substantially depleted. The relevant intervention is consequently not simply to maximize κ , but to reinforce collective autonomy together with sufficiently strong return processes ρ so that protection does not come at the cost of a broad hysteretic regime.

The parameters μ and σ play a different role. They do not shift λ_{SN} or λ_{TC} , but control the composition of the coupled population. At equilibrium, we have

$$\frac{D^*}{C^* + D^*} = \frac{\mu}{\mu + \sigma}, \quad (20)$$

so decreasing μ or increasing σ reduces the fraction of users in the dependent state without necessarily reducing overall LLM adoption. These parameters therefore describe a second class of interventions aimed not at preventing coupling itself, but at limiting its transition toward persistent cognitive substitution. Examples include metacognitive training, verification requirements, periodic unaided practice, task designs that preserve active reasoning, and mechanisms that facilitate recovery from dependency.

The model thus distinguishes two complementary intervention levels. Changes in λ , ρ , and κ reshape the population-level tipping landscape and determine whether bistability and hysteresis are possible, whereas changes in μ and σ primarily determine the cognitive burden associated with the coupled state. In this sense, immunization can act either by preventing a collective transition or by reducing the probability that ordinary LLM use develops into persistent dependency.

Thus immunization (as defined here) need not imply low overall LLM use. It can instead consist of shaping the coupling so that high adoption remains compatible with verification, active reasoning, autonomous alternatives, and recovery from substitutive dependence.

V. DISCUSSION

In 1960, Joseph Licklider anticipated a future in which the relation between humans and computers would move beyond simple tool use [107]:

“Man-computer symbiosis is probably not the ultimate paradigm for complex technological systems. It seems entirely possible that, in due course, electronic or chemical machines will outdo the human brain in most of the functions we now consider exclusively within its province.”

More than six decades later, LLMs made this quote especially concrete. Their importance lies not only in

Intervention	Main mathematical effect	Effect on tipping and cognitive state
Reduce propagation of substitutive/dependency-producing coupling	Decreases λ by limiting automatic adoption and social or institutional amplification of LLM use.	Direct. Prevents invasion for $\lambda < \lambda_{\text{TC}}$; after lock-in, recovery requires $\lambda < \lambda_{\text{SN}}$.
Preserve autonomous alternatives and routes back to unaided cognition	Increases ρ , the rate at which regular users return to uncoupled or weakly coupled cognition.	Raises both thresholds and reduces hysteresis. Bistability disappears for $\rho \geq \kappa$.
Strengthen collective autonomy	Increases κ , the strength of nonlinear collective reinforcement of autonomous cognition.	Raises resistance to invasion, but for $\kappa > \rho$ also widens the hysteretic interval.
Favor recovery over collective lock-in	Increases the ratio ρ/κ , strengthening individual routes back to autonomous cognition relative to collective reinforcement.	Moves the system toward $\rho = \kappa$, where the saddle-node and transcritical thresholds merge and the transition becomes continuous.
Prevent progression to dependency	Decreases μ , reducing transitions from regular use C to persistent dependency D .	No direct effect on λ_{SN} or λ_{TC} . Reduces $D^*/(C^* + D^*)$ and increases $\langle \Gamma \rangle$.
Promote recovery from dependency	Increases σ , shifting users from persistent dependency D back to regular use C .	No direct effect on the bifurcation thresholds. Reduces dependency and raises $\langle \Gamma \rangle$ at fixed U^* .

TABLE I. **Qualitative effects of cognitive immunization strategies.** The parameters λ , ρ , and κ modify the bifurcation structure, with $\lambda_{\text{SN}} = 2\sqrt{\kappa\rho}$ and $\lambda_{\text{TC}} = \rho + \kappa$. For $\kappa > \rho$, these thresholds delimit the bistable regime. By contrast, μ and σ do not move the tipping points in this minimal model, but determine the partition between regular and dependent users and therefore the cognitive competence $\langle \Gamma \rangle$.

what they can do, but in the new forms of coupling they create between human and artificial cognition. At one extreme, AI could become an “exocortex”: an external extension of cognition capable of searching large bodies of knowledge while leaving interpretation and high-level judgment to the human researcher [108]. Such systems could greatly accelerate the accumulation and transmission of cultural and scientific knowledge. More generally, humans have always externalized parts of cognition into language, writing, institutions, and technology. This externalization has been the engine of rapid and accelerated change driven by culture [109, 110]. AI represents a new step in this process, one in which external cognitive machinery becomes increasingly active and agent-like [111]. As suggested in previous studies, this might be a new class of *synthetic* evolutionary transition [112–115].

Our model suggests that the consequences of this transition cannot be understood from individual LLM use alone. Two social processes interact. First, new ways of using LLMs spread because people learn from other people, because workplaces and schools adopt them, and because successful practices are copied. In this sense, adoption has a contagion-like component. Second, autonomous cognition is itself socially sustained. Education and institutions do more than transmit information: they create environments in which reading, writing, reasoning, verification, and independent problem solving are repeatedly practiced and rewarded. When these practices are common, they reinforce each other.

The interaction between these two processes creates the central feedback in our model. As cognitive delegation becomes more widespread, the social environment

that supports autonomous reasoning can weaken; as that environment weakens, delegation becomes still easier and more attractive. Therefore, a gradual increase in LLM adoption can produce a disproportionate collective response. Beyond a critical point, the loss of autonomy becomes self-reinforcing and the population can move rapidly toward a state of much stronger cognitive offloading and lower cognitive competence. The important result is not that such a runaway must occur, but that it *can* occur under plausible forms of social learning and cooperative reinforcement. Moreover, once such a state has become established, simply returning conditions to where they were before the transition may not be sufficient to restore the previous state. Prevention can therefore be considerably easier than reversal.

This possibility should not obscure the substantial benefits of LLMs. The distinction is not between using and not using AI, but between forms of coupling that extend human competence and those that replace the cognitive operations through which competence is maintained. The evidence already points to both possibilities. Recent experiments also show that the direction of these effects strongly depends on the architecture of the human-AI coupling. In a randomized study, a guided “think first, ChatGPT later” protocol produced higher subsequent independent creativity than unrestricted use of the ChatGPT [116]. Moreover, generative AI can increase productivity and provide powerful cognitive assistance, but knowledge workers also report a reduced effort to think critically when confidence in AI is high [117]. In education, unrestricted access to generative AI can improve performance while the tool is available, yet

reduce subsequent unaided performance, while pedagogically constrained AI can substantially mitigate this effect [118, 119]. Related results indicate poorer comprehension or retention when students rely on LLMs without engaging in complementary cognitive activities such as note-taking [120]. In programming, novice users can also struggle to understand and critically evaluate AI-generated code, creating conditions for automation bias and superficial competence [121].

These findings reinforce the distinction between AI as *scaffolding* and AI as *substitution*. This distinction also changes how we should think about “cognitive immunization.” At the population level, immunization does not mean preventing contact with AI. It means preserving the practices and institutions that keep human cognition active: unaided problem solving, verification, critical discussion, periods of deliberate disengagement, maintenance of non-AI skills, and educational designs in which the model supports rather than completes the cognitive task. However, once persistent dependency has developed, the problem can no longer be addressed by educational design alone. Emerging work on problematic LLM use points to psychological and behavioral mechanisms related to loss of control, emotional regulation, cognitive biases, and habitual reliance, suggesting that behavioral self-regulation and, in more severe cases, psychological interventions such as cognitive-behavioral therapy may become relevant [122]. Thus, the analogy with immunization extends naturally from prevention at the cultural and institutional level to recovery at the individual level.

Several extensions could bring this minimal model closer to the complexity of real human-LLM interactions. These include replacing the population-level description with heterogeneous interacting agents [22], allowing cognitive autonomy to vary continuously rather than dividing users into sharply separated classes, and introducing feedbacks through which cognitive change modifies subsequent adoption and dependence. Such a formulation would also soften the boundaries between humans, LLMs, and increasingly hybrid human-machine systems [123], which may occupy a broad continuum of cognitive organizations rather than discrete categories. More generally, if agents themselves can be understood as persistent patterns of information and activity distributed on biological, artificial, or hybrid substrates [111, 124, 125], the relevant evolving entities may not always be conventional individuals. What spreads could instead be patterns of cognitive organization, including perspectives that shape how agents represent the world, themselves, and possible actions. LLMs would then be more than passive tools for cognitive offloading: by participating in these distributed patterns, they could progressively reshape the cognitive structures through which users interpret and act on the world. From this perspective, the “cognitive virus” is a persistent technological lineage embedded in a larger ecology of cultural replicators, human host states, institutions, and technical infrastructure.

A further limitation is that human behavior is not yet explicitly coupled to the spreading process. Such feedbacks can qualitatively reshape epidemic thresholds [126–130], and can even generate higher-order critical phenomena [131–133]. A useful conceptual framework for extending this picture is provided by the emerging field of *Diverse Intelligence*, which treats intelligence and agency not as properties restricted to brains or particular biological substrates, but as graded capacities of systems operating across multiple spatial, temporal, and organizational scales [134, 135]. In particular, this perspective emphasizes the ability of biological, artificial, and hybrid agents [123] to navigate problem spaces and maintain goal-directed organization, providing a natural language to describe coupled human-AI systems in which boundaries and degrees of agency may themselves change over time. In addition, multilayer-network theory offers a formal framework for representing the interacting behavioral, technological, and social processes through which such feedbacks propagate [136, 137]. Together, these approaches suggest a systematic route beyond the present mean-field description.

Future models could therefore treat autonomy, dependence, and agency as continuous and coevolving properties of distributed human-machine patterns, rather than as fixed compartments, and ask under what conditions transient interactions become self-maintaining forms of cognitive organization.

ACKNOWLEDGMENTS

R.S. has been supported by an AGAUR 2021 SGR 0075 grant, by AEI-PID2023-152129NB-I00 grant funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU, and the Santa Fe Institute. S.F.E. has been supported by grants PID2025-169610NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by “ERDF a way of making Europe”, CIPROM/2022/59 funded by Generalitat Valenciana, and the Santa Fe Institute. L.F.S. has been supported by the Occident Foundation (grant FJSCNB-2022-12-B) and by the Spanish State Research Agency, AEI, through grant PID2023-153225NA-I00 funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU. L.F.S. conducted this research while visiting the Okinawa Institute of Science and Technology (OIST) through the Theoretical Sciences Visiting Program (TSVP).

REFERENCES

- [1] E. Szathmáry and J. Maynard Smith, *Nature* **374**, 227 (1995).
- [2] M. A. Nowak and D. C. Krakauer, *Proceedings of the National Academy of Sciences* **96**, 8028 (1999).
- [3] M. Tomasello, *The Cultural Origins of Human Cognition* (Harvard University Press, Cambridge, MA, 1999).

- [4] R. Boyd and P. J. Richerson, *Culture and the Evolutionary Process* (University of Chicago Press, Chicago, 1985).
- [5] R. V. Solé, B. Corominas-Murtra, and J. Fortuny, *Journal of the Royal Society Interface* **7**, 1647 (2010).
- [6] M. Tomasello, *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Harvard University Press, Cambridge, MA, 2003).
- [7] W. S. Burroughs, *The Electronic Revolution* (Expanded Media Editions, 1970).
- [8] K. Aoki, *University of Miami Law Review* **53**, 961 (1999).
- [9] F. Cohen, *Computers & Security* **6**, 22 (1987).
- [10] M. W. Feldman and K. N. Laland, *Trends in Ecology & Evolution* **11**, 453 (1996).
- [11] H. Gintis, *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 878 (2011).
- [12] H. Whitehead, K. N. Laland, L. Rendell, R. Thorogood, and A. Whiten, *Nature Communications* **10**, 2405 (2019).
- [13] R. Dawkins, *The Selfish Gene* (Oxford University Press, Oxford, 1976).
- [14] S. Blackmore, *The Meme Machine* (Oxford University Press, Oxford, 2000).
- [15] D. C. Dennett, *From bacteria to Bach and back: The evolution of minds* (WW Norton & Company, 2017).
- [16] R. Dawkins, in *Dennett and His Critics: Demystifying Mind*, edited by B. Dahlbom (Blackwell, Cambridge, MA, 1993) pp. 13–27.
- [17] R. Brodie, *Virus of the mind: The new science of the meme* (Hay House, Inc, 2009).
- [18] A. Clark and D. Chalmers, *Analysis* **58**, 7 (1998).
- [19] D. A. Norman, in *Designing Interaction: Psychology at the Human-Computer Interface*, edited by J. M. Carroll (Cambridge University Press, Cambridge, 1991) pp. 17–38.
- [20] E. Hutchins, *Cognitive Science* **19**, 265 (1995).
- [21] D. C. Krakauer, *Nautilus* (2016), issue 40.
- [22] D. C. Krakauer, arXiv preprint arXiv:2607.18460 (2026).
- [23] M. Fasoli, *Review of Philosophy and Psychology* **9**, 671 (2018).
- [24] A. Clark, in *Language and Thought: Interdisciplinary Themes*, edited by P. Carruthers and J. Boucher (Cambridge University Press, Cambridge, 1998) pp. 162–183.
- [25] J. Winawer, N. Witthoft, M. C. Frank, L. Wu, A. R. Wade, and L. Boroditsky, *Proceedings of the National Academy of Sciences* **104**, 7780 (2007).
- [26] G. Lupyan and E. J. Ward, *Proceedings of the National Academy of Sciences* **110**, 14196 (2013).
- [27] C. M. Fausey, B. L. Long, A. Inamori, and L. Boroditsky, *Frontiers in Psychology* **1**, 162 (2010).
- [28] P. H. Thibodeau and L. Boroditsky, *PLOS ONE* **6**, e16782 (2011).
- [29] S. Kirby, H. Cornish, and K. Smith, *Proceedings of the National Academy of Sciences* **105**, 10681 (2008).
- [30] M. H. Christiansen and N. Chater, *Creating language: Integrating evolution, acquisition, and processing* (The MIT press, 2016).
- [31] S. A. Gelman and S. O. Roberts, *Proceedings of the National Academy of Sciences* **114**, 7900 (2017).
- [32] D. Crystal, *Language and the Internet*, 2nd ed. (Cambridge University Press, Cambridge, 2006).
- [33] N. S. Baron, *Always On: Language in an Online and Mobile World* (Oxford University Press, Oxford, 2008).
- [34] R. Ling and N. S. Baron, *Journal of Language and Social Psychology* **26**, 291 (2007).
- [35] F. Lyddy, F. Farina, J. Hanney, L. Farrell, and N. K. O'Neill, *Journal of Computer-Mediated Communication* **19**, 546 (2014).
- [36] I. Ruiz-Trillo, E. Casacuberta, N. H. Brown, and R. Solé, *The EMBO Journal* **45**, 1865 (2026).
- [37] d. boyd, in *A Networked Self: Identity, Community, and Culture on Social Network Sites*, edited by Z. Papacharissi (Routledge, New York, 2010) pp. 39–58.
- [38] J. Androutsopoulos, *Discourse, Context & Media* **4-5**, 62 (2014).
- [39] G. Kress, *Multimodality: A Social Semiotic Approach to Contemporary Communication* (Routledge, London, 2010).
- [40] M. Zappavigna, *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web* (Continuum, London, 2012).
- [41] F. Cohen, *Computers & Security* **8**, 297 (1989).
- [42] S. Cass, *IEEE Spectrum* **38**, 56 (2001).
- [43] J. Parikka, *Digital contagions: A media archaeology of computer viruses*, Vol. 44 (Peter Lang, 2007).
- [44] J. E. Sammet, *Communications of the ACM* **15**, 601 (1972).
- [45] C. Nachenberg, *Communications of the ACM* **40**, 46 (1997).
- [46] E. H. Spafford, *Artificial life* **1**, 249 (1994).
- [47] R. Pastor-Satorras and A. Vespignani, *Physical review letters* **86**, 3200 (2001).
- [48] A. L. Lloyd and R. M. May, *Science* **292**, 1316 (2001).
- [49] M. J. Keeling and K. T. Eames, *Journal of the royal society interface* **2**, 295 (2005).
- [50] S. Forrest, S. A. Hofmeyr, and A. Somayaji, *Communications of the ACM* **40**, 88 (1997).
- [51] D. Dasgupta and F. Nino, *Immunological computation: theory and applications* (CRC press, 2008).
- [52] S. Valverde and R. V. Solé, *Journal of The Royal Society Interface* **12**, 20150249 (2015).
- [53] S. Valverde and R. V. Solé, *Human biology* **87**, 224 (2015).
- [54] A. J. Hey and G. Pápay, *The computing universe: a journey through a revolution* (Cambridge University Press, 2015).
- [55] M. Gerlich, *Societies* **15**, 6 (2025).
- [56] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).
- [57] M. Karshenas and P. L. Stoneman, *The RAND Journal of Economics* **24**, 503 (1993).
- [58] J. Dahlke, M. Beck, J. Kinne, D. Lenz, R. Dehghan, M. Wörter, and B. Ebersberger, *Research Policy* **53**, 104917 (2024).
- [59] M. M. Lehman, *Information Processing & Management* **20**, 19 (1984).
- [60] M. M. Lehman, *Information and Software Technology* **38**, 681 (1996).
- [61] B. Fitzgerald and K.-J. Stol, *Journal of Systems and Software* **123**, 176 (2017).
- [62] R. Ros, E. Bjarnason, and P. Runeson, *Empirical Software Engineering* **29**, 21 (2024).
- [63] L. M. Márquez, R. S. Redman, R. J. Rodriguez, and M. J. Roossinck, *Science* **315**, 513 (2007).
- [64] F. Ryan, *Volution* (HarperCollins, 2009).

- [65] M. J. Roossinck, *Nature Reviews Microbiology* **9**, 99 (2011).
- [66] J. J. Barr, R. Auro, M. Furlan, K. L. Whiteson, M. L. Erb, J. Pogliano, A. Stotland, R. Wolkowicz, A. S. Cutting, K. S. Doran, P. Salamon, M. Youle, and F. Rohwer, *Proceedings of the National Academy of Sciences* **110**, 10771 (2013).
- [67] M. J. Roossinck and E. R. Bazán, *Annual Review of Virology* **4**, 123 (2017).
- [68] S. Noy and W. Zhang, *Science* **381**, 187 (2023).
- [69] E. Brynjolfsson, D. Li, and L. R. Raymond, *The Quarterly Journal of Economics* **140**, 889 (2025).
- [70] E. F. Risko and S. J. Gilbert, *Trends in cognitive sciences* **20**, 676 (2016).
- [71] A. R. Doshi and O. P. Hauser, *Science Advances* **10**, eadn5290 (2024).
- [72] F. M. Bass, *Management Science* **15**, 215 (1969).
- [73] E. White and C. Comiskey, *Mathematical biosciences* **208**, 312 (2007).
- [74] N. A. Battista, L. B. Pearcy, and W. C. Strickland, *Bulletin of mathematical biology* **81**, 2258 (2019).
- [75] M. W. Van den Ende, S. Epskamp, M. H. Lees, H. L. van der Maas, R. W. Wiers, and P. M. Sloot, *Addictive behaviors* **127**, 107201 (2022).
- [76] C. Chou, L. Condron, and J. C. Belland, *Educational psychology review* **17**, 363 (2005).
- [77] D. J. Kuss, M. D. Griffiths, L. Karila, and J. Billieux, *Current pharmaceutical design* **20**, 4026 (2014).
- [78] H. T. Alemneh and N. Y. Alemu, *Infectious Disease Modelling* **6**, 405 (2021).
- [79] F. Courchamp, L. Berec, and J. Gascoigne, *Allee effects in ecology and conservation* (OUP Oxford, 2008).
- [80] R. Solé, *Phase transitions* (Princeton University Press, 2011).
- [81] M. De Domenico, *Reports on Progress in Physics* **89**, 014601 (2026).
- [82] M. E. Newman, *Physical review E* **66**, 016128 (2002).
- [83] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Reviews of modern physics* **87**, 925 (2015).
- [84] R. M. May and A. L. Lloyd, *Physical Review E* **64**, 066112 (2001).
- [85] R. Pastor-Satorras and A. Vespignani, *Physical review E* **65**, 036104 (2002).
- [86] R. Cohen, S. Havlin, and D. Ben-Avraham, *Physical review letters* **91**, 247901 (2003).
- [87] M. Boguná and R. Pastor-Satorras, *Physical Review E* **66**, 047104 (2002).
- [88] M. E. Newman, *Physical review letters* **89**, 208701 (2002).
- [89] M. E. Newman, *Physical Review E* **68**, 026121 (2003).
- [90] J. C. Miller, *Journal of the Royal Society Interface* **6**, 1121 (2009).
- [91] M. Salathé and J. H. Jones, *PLoS computational biology* **6**, e1000736 (2010).
- [92] N. Zomer and M. De Domenico, *npj Artificial Intelligence* **2**, 10.1038/s44387-026-00091-5 (2026).
- [93] T. Gross, C. J. D. D’Lima, and B. Blasius, *Physical review letters* **96**, 208701 (2006).
- [94] V. Colizza and A. Vespignani, *Physical review letters* **99**, 148701 (2007).
- [95] A. Brechtel, P. Gramlich, D. Ritterskamp, B. Drossel, and T. Gross, *Physical Review E* **97**, 032307 (2018).
- [96] J. Nauta and M. De Domenico, *Physical Review Research* **6**, 043055 (2024).
- [97] S. Kéfi, M. Rietkerk, M. van Baalen, and M. Loreau, *Theoretical population biology* **71**, 367 (2007).
- [98] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara, *Nature* **461**, 53 (2009).
- [99] M. Granovetter, *American Journal of Sociology* **83**, 1420 (1978).
- [100] P. Ball, *Critical mass: How one thing leads to another* (Macmillan, 2004).
- [101] C. Castellano, S. Fortunato, and V. Loreto, *Reviews of modern physics* **81**, 591 (2009).
- [102] D. Centola, J. Becker, D. Brackbill, and A. Baronchelli, *Science* **360**, 1116 (2018).
- [103] A. F. Ashery, L. M. Aiello, and A. Baronchelli, *Science Advances* **11**, eadu9368 (2025).
- [104] B. Jose, J. Cherian, A. M. Verghis, S. M. Varghise, S. Mumthas, and S. Joseph, *Frontiers in Psychology* **16**, 1550621 (2025).
- [105] S. Dehaene, L. Cohen, J. Morais, and R. Kolinsky, *Nature Reviews Neuroscience* **16**, 234 (2015).
- [106] N. Kosmyna, E. Hauptmann, Y. T. Yuan, J. Situ, X.-H. Liao, A. V. Beresnitzky, I. Braunstein, and P. Maes, *arXiv preprint arXiv:2506.08872* (2025).
- [107] J. C. R. Licklider, *IRE Transactions on Human Factors in Electronics* **HFE-1**, 4 (1960).
- [108] K. G. Yager, *Digital Discovery* **3**, 1933 (2024).
- [109] M. Muthukrishna, M. Doebeli, M. Chudek, and J. Henrich, *PLoS computational biology* **14**, e1006504 (2018).
- [110] C. Perreault, *Proceedings of the National Academy of Sciences* **123**, e2523038123 (2026).
- [111] G. Ruffini and F. Castaldo, *Preprints* 10.20944/preprints202607.0418.v2 (2026), version 2, posted 14 July 2026.
- [112] R. Sole, *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150438 (2016).
- [113] P. B. Rainey, *Philosophical Transactions of the Royal Society B: Biological Sciences* **378**, 20210408 (2023).
- [114] P. B. Rainey and M. E. Hochberg, *Proceedings of the National Academy of Sciences* **122**, e2509122122 (2025).
- [115] V. Müller, L. Steels, and E. Szathmáry, *Proceedings of the National Academy of Sciences* **123**, e2527700123 (2026).
- [116] S. S. H. Wong and S. X. Qiu, *Educational Psychology Review* **38**, 45 (2026).
- [117] H.-P. Lee, A. Sarkar, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, and N. Wilson, in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, New York, NY, USA, 2025) pp. 1–22.
- [118] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakçı, and R. Mariman, *Proceedings of the National Academy of Sciences* **122**, e2422633122 (2025).
- [119] F. Huettig and M. H. Christiansen, *Cognitive Science* **48**, e13487 (2024).
- [120] P. Kreijkes, V. Kewenig, M. Kuvalja, M. Lee, J. M. Hofman, S. Vitello, A. Sellen, S. Rintel, D. G. Goldstein, D. Rothschild, L. Tankelevitch, and T. Oates, *Computers & Education* **243**, 105514 (2026).
- [121] Y. Zi, L. Li, A. Guha, C. J. Anderson, and M. Q. Feldman, in *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering* (Association for Computing Machinery, 2025) pp. 1479–1488.

- [122] H.-Y. Liao, C.-H. Ko, and C.-F. Yen, *Biomedical Journal* **49**, 100998 (2026).
- [123] R. Solé, L. F. Seoane, J. Pla-Mauri, M. T. Bennett, M. E. Hochberg, and M. Levin, *arXiv preprint arXiv:2601.12837* (2026).
- [124] M. Levin, *Frontiers in systems neuroscience* **16**, 768201 (2022).
- [125] C. Fields and M. Levin, *Physics of Life Reviews* **52**, 256 (2025).
- [126] E. P. Fenichel, C. Castillo-Chavez, M. G. Ceddia, G. Chowell, P. A. G. Parra, G. J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, *et al.*, *Proceedings of the National Academy of Sciences* **108**, 6306 (2011).
- [127] M. C. Bootsma and N. M. Ferguson, *Proceedings of the National Academy of Sciences* **104**, 7588 (2007).
- [128] S. Funk, M. Salathé, and V. A. Jansen, *Journal of the Royal Society Interface* **7**, 1247 (2010).
- [129] S. Funk, E. Gilad, C. Watkins, and V. A. Jansen, *Proceedings of the National Academy of Sciences* **106**, 6872 (2009).
- [130] P. Bosetti, P. Poletti, M. Stella, B. Lepri, S. Merler, and M. De Domenico, *Proceedings of the National Academy of Sciences* **117**, 30118 (2020).
- [131] C. Granell, S. Gómez, and A. Arenas, *Physical review letters* **111**, 128701 (2013).
- [132] A. Lima, M. De Domenico, V. Pejovic, and M. Musolesi, *Scientific reports* **5**, 10650 (2015).
- [133] D. Maniscalco, T. Bertola, V. d’Andrea, and M. De Domenico, *Chaos, Solitons & Fractals* **198**, 116501 (2025).
- [134] R. Chis-Ciure and M. Levin, *Synthese* **206**, 10.1007/s11229-025-05319-6 (2025), article 257.
- [135] M. Levin, *Advanced Intelligent Systems* **7**, 70130 (2025).
- [136] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, *Nature Physics* **12**, 901 (2016).
- [137] M. De Domenico, *Nature Physics* **19**, 1247 (2023).