

RESEARCH ARTICLE

Difficult control is related to instability in biologically inspired Boolean networks

Bryan C. Daniels ^{*}, Enrico Borriello ^{*}

School of Complex Adaptive Systems, Arizona State University, Tempe, Arizona, United States of America

^{*} bryan.daniels.1@asu.edu (BCD); enrico.borriello@asu.edu (EB)

Abstract

Previous work in Boolean dynamical networks has suggested that the number of components that must be controlled to select an existing attractor is typically set by the number of attractors admitted by the dynamics, with no dependence on the size of the network. Here we study the rare cases of networks that defy this expectation, with attractors that require controlling most nodes. We find empirically that unstable fixed points are the primary recurring characteristic of networks that prove more difficult to control. We describe an efficient way to identify unstable fixed points and show that, in both existing biological models and ensembles of random dynamics, we can better explain the variance of control kernel sizes by incorporating the prevalence of unstable fixed points. In the end, the association of these outliers with dynamics that are unstable to small perturbations reveals them as artifacts of deterministic models, making them less biologically relevant and reinforcing the generality of easy controllability in biological networks.

 OPEN ACCESS

Citation: Daniels BC, Borriello E (2025) Difficult control is related to instability in biologically inspired Boolean networks. *PLOS Complex Syst* 2(1): e0000025. <https://doi.org/10.1371/journal.pcsy.0000025>

Editor: Luis M. Rocha, Binghamton University, UNITED STATES OF AMERICA

Received: March 24, 2024

Accepted: October 16, 2024

Published: January 3, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcsy.0000025>

Copyright: © 2025 Daniels, Borriello. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data underlying these findings are fully available and public through Zenodo. The python code used to recreate the analysis and figures, as well as CSV files containing the relevant data, are available as a Zenodo

Author summary

What sets how easily a living system can be controlled? Such a question can be operationalized in terms of the number of system components that need to be forced to guarantee a desired outcome. Previous results in Boolean networks have suggested that control does not intrinsically become more difficult in larger networks, but instead depends mostly on the number of distinct long-term dynamics exhibited by the system. Yet there are exceptions to this rule, cases in which most or all nodes in a network must be controlled to achieve certain end states. Here, we study these cases in detail and show that they are related to instability. We view our results as encouraging for the hypothesis that even large biological networks may be typically easy to control.

1 Introduction

The formulation of an exhaustive theory for controlling complex biological systems, and gene regulatory networks (GRNs) more specifically, remains a key objective in systems biology [1, 2]. And while optimal control theory offers a broad set of tools for various applications in mathematics and engineering, its traditional assumptions often require modification for

repository under the accession code [10.5281/zenodo.13819680](https://doi.org/10.5281/zenodo.13819680) [<https://doi.org/10.5281/zenodo.13819680>].

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

biological contexts. Biological processes are typically inherently nonlinear [3], and the focus is frequently restricted to coarse-grained output states [4], rather than driving the system to an arbitrary state [5]. There are a number of approaches to quantifying control in biology, which vary in their basic assumptions. States can be modeled as continuous or discrete, dynamics can be assumed to be stochastic or deterministic, control inputs can be static or dynamic, and the goal of control can be to reach arbitrary states or to reach particular phenotypes (attractors). Much work in biological control has focused on phenotypic control toward specific coarse-grained output states [4], as it is arguably more relevant to biological function than driving a system into any arbitrary state [5]. Some work inspired by traditional control theory has focused on control toward arbitrary states, either in continuous [6] or discrete state space [7, 8], though this typically leads to control strategies that are significantly more difficult to implement than what is necessary to select a single desired output. The simplicity of discrete state spaces (and the inability to constrain parameters in continuous formulations) has guided many to take a Boolean approach. The typical interventions available in experiments has led to a focus on control inputs that are largely static, either applied permanently as in a gene knock-out or temporarily over some timescale [9–13], though some approaches allow for arbitrary dynamics of inputs [7, 11]. Finally, for simplicity and mathematical tractability, the updating functions of Boolean network models are often defined deterministically. A mild form of stochasticity is sometimes incorporated in defining the order in which nodes are updated across time (e.g., [9–11, 13]), or noise can be modeled in the form of bit-flips or probabilistic transitions [14].

We focus here on discrete, Boolean, deterministic, static, phenotypic control. In this context, an important concept for regulating network dynamics is the *control kernel* (CK). First proposed in [15], the CK refers to a set of genes of minimal size whose external manipulation is enough to guide the network towards a target steady-state gene activation pattern, i.e. an attractor of the dynamical model of the GRN. The CK is defined such that it can force a particular attractor state regardless of the initial condition. This matches with the assumption that biological function is mostly captured by attractors, in that they represent phenotypically and functionally distinct cell types [16].

Several related approaches have been proposed for the study of the optimal control of deterministic Boolean networks. Stable motif analysis defines control sets by detecting positive circuits within the network that can sustain trap spaces in the dynamics [17]. Feedback vertex sets offer an efficient upper bound on the size of minimal control sets in a way that does not depend on the specific dynamics governing each node [5, 18, 19].

A simple heuristic argument suggests that the size of CKs may scale approximately with the logarithm of the number of original attractors. Under the simplifying assumption that state transitions are all equally likely, forcing a single node into a specific state would generally halve the number of attainable states. Therefore, by controlling c nodes, we would expect to reduce 2^c possible attractors down to a single attractor, suggesting the logarithmic scaling of CK size. However, the task of identifying the minimal set of controlling nodes is nontrivial and has been proven NP-hard by Akutsu *et al.* [20], and counterexamples can be constructed that require much larger CKs.

In Ref. [21], we demonstrated that dynamic Boolean networks do often have CK sizes that scale logarithmically with the number of attractors. We illustrated this on a large database of experimentally derived biological networks, as well as ensembles of random networks. These previous findings revealed that control toward existing attractors does not inherently become more difficult as the size of the network increases. Instead, the number of attractor states generated by the dynamics is a much better predictor of the difficulty of control than the network size. Our result could then be summarized in the statement that the average size of the control

kernel of a dynamical Boolean network, $\langle |CK| \rangle$, typically scales logarithmically with the number of its attractors, r :

$$\langle |CK| \rangle \simeq \log_2 r. \quad (1)$$

In Ref. [21], we reinterpreted this result in terms of the conjectured expectation value of the witness set in computational learning theory [22]. More importantly, we view it as one of the most significant theoretical and empirical justifications to date for the feasibility of genetic reprogramming [23].

Unbeknownst to us, research conducted by Akutsu's group and published earlier than our work had reached a very similar conclusion and independently identified the logarithmic scaling described, even if under somewhat different assumptions and on more abstract models [24]. A number of clever simplifications make the mathematics more manageable and allow for the direct derivation of logarithmic scaling as in Eq 1. The primary difference between the control defined by Hou et al. and our study is the duration of the pinning. Hou et al. assume that the pinning, which constitutes the control signal, lasts for a single timestep and is then removed, whereas we assume that the pinning is permanent. Our choice of permanent pinning is influenced by our focus on biological networks, where long-timescale perturbations, such as those observed in gene overexpression, knockdown, and knockout experiments, have practical significance. Another important difference is that Hou et al. assume that the basins of all attractors are of equal size. Due to the large variability in the basin sizes of the experimentally motivated models we examine, our current work aims to shed more light on the still hazy relationship between basin size and controllability.

In our previous study, some networks were much more difficult to control than predicted by a logarithmic scaling law (roughly 2% of tested networks). A few networks had significant numbers of attractors whose control required pinning nearly all nodes in the network. The presence of such networks, along with the unaccounted reasons for their difference from more easily controllable networks, could undermine one's confidence in predicting the required level of control based solely on knowledge of a network's attractors, as well as the hypothesis that a network's size is typically irrelevant to the difficulty of control.

Our focus in this work is to investigate the factors that contribute to the increased difficulty of controlling such networks. Our goal is to find approximations to the mean control kernel size that are simple to compute and do not rely on solving the full dynamics, which becomes intractable in moderately sized networks.

As a first step, we expand our database of biological networks by incorporating additional networks collected in Ref. [25], representing the most extensive collection of Boolean models for biological networks at the time of writing this manuscript. None of the new networks exhibit a significant deviation from our original predictions, leaving us with the task of explaining only the outliers previously identified.

A direct examination of these outlier networks highlights the presence of isolated fixed points, i.e., fixed points of the dynamics that do not attract any other state. We establish a strong correlation between this characteristic and the difficulty of control, and we use this as the rationale for a correction to our original scaling law (Eq 1). The revised scaling law relies not only on the knowledge of the number of attractors but also on the number of isolated fixed points among them. Interestingly, the latter can be easily evaluated even for networks where not all basin sizes are known, particularly in cases where the identity of all attractors can be obtained by exploiting modularity in the network topology. We test whether our new empirical formula outperforms our previous scaling law, and demonstrate that it accounts for the majority of the remaining variance not captured by the original scaling.

In the next section (*Results*), we describe our extended analysis and provide the correction to our empirical formula.

In the *Methods* section, we show how we leverage the modular structure of biological networks to identify the complete spectrum of their attractor states, and we demonstrate how to exploit the sparse connectivity of these networks to identify all isolated fixed-points by explicitly determining their backward reachable sets, i.e., the sets of their pre-images.

Finally, in the *Discussion* section, we comment on the biological relevance of these isolated fixed-points.

2 Results

To explore the drivers of unusually large control kernels, we first ask: What is the maximum possible control kernel size for a given network? Given our definition of control, fixed-point attractors always have a control kernel, as control can be attained by pinning all nodes in the network. Note that cyclic attractors are sometimes not able to be forced with static control because we are limited to pinning non-cycling nodes—in these cases, we say that a control kernel does not exist. This sets the simplest upper bound on the size of control kernels for controllable attractors of $|\text{CK}| \leq n$. A more economical approach avoids pinning nodes in any peripheral trees that are wholly dependent on nodes within closed feedback loops in the regulatory network topology. We will refer to the network's core as the set of nodes that exclude any peripheral trees, with a number of core nodes n_c . As control kernels are defined as sets of controlling nodes of minimal size, we can ensure a tighter bound of $|\text{CK}| \leq n_c$.

We might naively expect that it would be more difficult to steer the system toward attractors with a smaller than average basin size when initializing the system in a random state. Therefore, one might expect extreme control kernel sizes for attractors with very small basins of attraction (defined as the states that lead to each attractor under the dynamics). While basin sizes are not generally easy to compute, it is relatively simple to highlight the most extreme scenarios when an attractor is an “isolated” fixed point, i.e., with a basin size of 1 (see *Methods*).

Both isolated fixed points and maximal control kernel sizes ($|\text{CK}| = n_c$) are connected to concepts of instability (examples shown in *Fig 1* and summarized in *Fig 2*). First, fixed points with maximal control kernels are locally unstable, in the sense that such a fixed point must be unstable to individual bit flips. That is, each state at Hamming distance 1 from the fixed point must be in the basin of a different attractor (otherwise one of the nodes could be left unpinned and the system would still return to the fixed point). Note that local instability does not necessarily imply that no other states transition into the fixed point under the dynamics (a counterexample is shown in the last row of *Fig 1*). Also, while having a maximal control kernel implies local instability, local instability does not in itself guarantee a maximal control kernel. Second, isolated fixed points are globally unstable, in the sense that no other states transition into them. (This is distinct from other notions of instability in the literature on Boolean networks, such as instability of cycles to delays in updating [26].) Though a fixed point being globally unstable does not directly imply that it has a maximal control kernel, we might expect some correlation between instability and difficulty of control.

In *Fig 3*, we compare average control kernel sizes in networks that have isolated (globally unstable) fixed points to the distribution over all networks (A and B) and control kernel sizes for individual isolated fixed points compared to the distribution over all attractors (C and D). The control kernel sizes of isolated fixed points are often close to n_c , matching with our intuition. Some of the biological networks have no feedback loops and are therefore controlled entirely by their input nodes, which trivially creates isolated fixed points with $|\text{CK}| = n_c$.

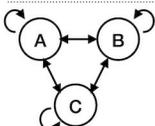
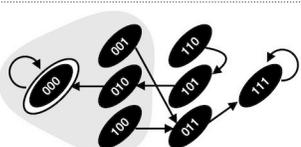
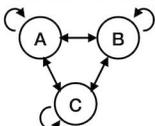
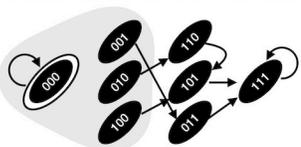
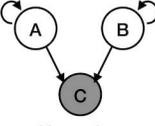
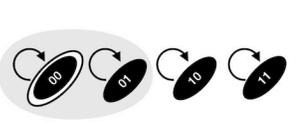
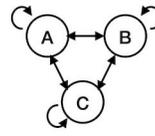
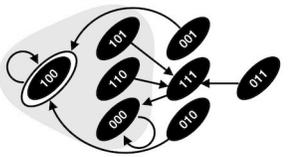
Example interaction network	Example state transition graph	Control kernel for highlighted state	Does the network have exceptional states?	Observed networks with these properties
 Has cycles		{C}	✗ Max CK	Biological 84% Random 79%
 Has cycles		{A,B,C}	✓ Max CK ✓ Locally unstable ✓ Globally unstable	Biological 6% Random 21%
 No cycles		{A,B}	✓ Max CK ✓ Locally unstable ✓ Globally unstable	Biological 10% Random 0%
 Has cycles		{A,B,C}	✓ Max CK ✓ Locally unstable ✗ Globally unstable	Biological 0% Random 0%

Fig 1. Examples of networks with and without exceptional states. Networks that have control kernels of maximal size n_c (“Max CK”) are related to those that have locally and globally unstable states. Empirically, most networks do not have control kernels of maximal size (first row). All observed networks that we identify as being unusually difficult to control contain both fixed points with maximal CK size and globally unstable fixed points (second row). Some of the biological networks have no cycles in their interaction networks, which trivially creates 2^m fixed points with $|CK| = m = n_c$; since they follow Eq 1, we do not consider these networks to be unusually difficult to control (third row). Finally, it is possible to construct networks that contain fixed points with maximal CK size but no globally unstable fixed points, but we find no examples of this case in the sampled networks (fourth row). Globally unstable fixed points have an incoming edge only from themselves in the state transition graph. Locally unstable fixed points can have other incoming edges, but none from nearest Hamming neighbors (gray region). Fixed points with maximal CK size must be locally unstable.

<https://doi.org/10.1371/journal.pcsy.0000025.g001>

(hatched areas in Fig 3 and third row in Fig 1). Even aside from this effect, isolated fixed points tend to have very large control kernels.

Thus, though global instability does not always imply difficult control, the correlation is nonetheless striking in the ensembles that we study. In fact, we find in our network ensembles

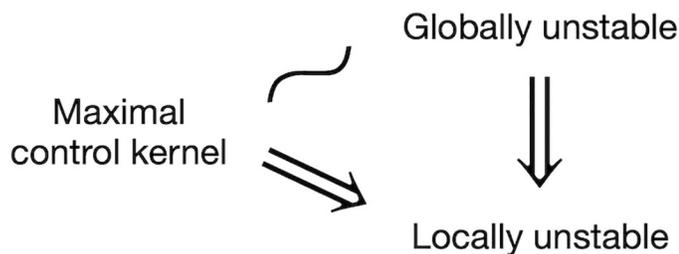


Fig 2. Summary of the logical relationships between the stability of a fixed point and whether it has a control kernel of maximal size. A fixed point with maximal control kernel size ($|CK| = n_c$) is always locally unstable, as is one that is globally unstable. Global instability is correlated with maximal control kernel size in the ensembles we study, but neither implies the other in general. Here local instability means that all states at Hamming distance 1 are in the basins of other attractors, and global instability means that no other states lead to the fixed point (basin size = 1).

<https://doi.org/10.1371/journal.pcsy.0000025.g002>

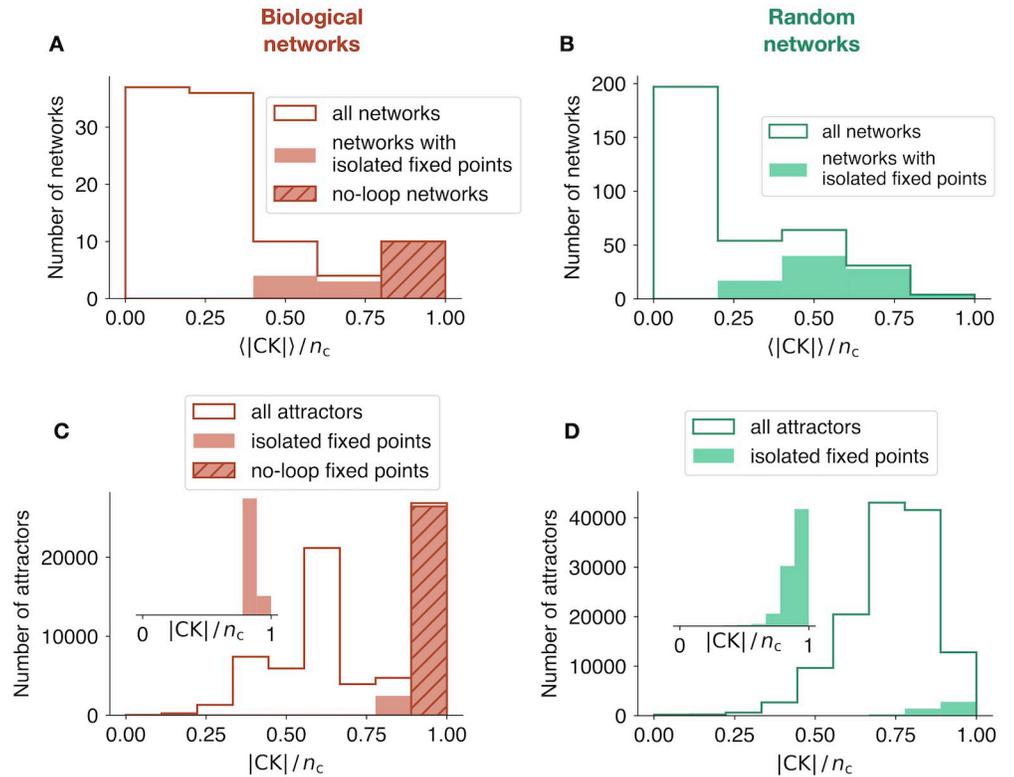


Fig 3. Isolated fixed points are related to large control kernel sizes. Both when averaged across attractors within networks (A and B) and at the level of individual attractors (C and D), isolated fixed points (those with basin size = 1; shaded bars) produce control kernel sizes $|CK|$ that are large relative to the total number of core nodes n_c . This pattern holds in both the biological (A and C) and random (B and D) network ensembles that we study. Some of the biological networks have no feedback loops, which trivially creates isolated fixed points with $|CK| = n_c$ (hatched bars in A and C). Insets in C and D highlight the distributions of control kernel sizes for only the isolated fixed points in networks that do contain feedback loops.

<https://doi.org/10.1371/journal.pcsy.0000025.g003>

that fixed points with maximal control kernels occur only in networks that have at least one globally unstable fixed point (Table 1).

At the level of individual networks and attractors there are exceptions to the rule that globally unstable fixed points tend to be difficult to control. See S3 and S4 Figs for histograms of control kernel sizes within each network that has at least one globally unstable fixed point. (In one network from the random ensemble, a globally unstable fixed point in fact has the smallest control kernel of all the network’s attractors.) Yet these exceptions are rare.

Table 1. Summary of the analyzed networks. “Total” includes all networks for which we can compute attractors and control kernels exactly. “Controllable” includes all networks for which static control kernels exist for a non-negligible set of attractors (here set as a threshold that the basins of uncontrollable cycles must make up less than 99% of the state space). “With Isolated” includes networks that have at least one isolated fixed point. “With Max CK” includes networks that have at least one fixed point with $|CK| = n_c$. “With Both” includes networks that have nonzero numbers of both isolated fixed points and maximal control kernels. (Note: 10 of the biological networks have no loops, which implies that all of their attractors are isolated and control kernels are of maximal size).

	Total	Controllable	With Isolated	With Max CK	With Both
Biological networks	104	97	17	16	16
Random networks	371	346	87	73	73

<https://doi.org/10.1371/journal.pcsy.0000025.t001>

Given that it is relatively fast to determine whether a given fixed point is globally unstable (see [Methods](#)), and these are also the ones that tend to contribute most to changes in $\langle |CK| \rangle$, we propose using their number to get a computationally simpler estimate of control kernel sizes. If we assume that isolated fixed points are the only ones contributing to the difference from [Eq 1](#) and that they have the maximum possible size n_c , we get the approximation

$$\langle |CK| \rangle \simeq \frac{r-s}{r} \log_2 r + \frac{s}{r} n_c, \tag{2}$$

where s is the number of isolated fixed-points. As shown in [Fig 4](#), this simple estimate does quite well in most cases, outperforming [Eq 1](#) for almost all networks we tested. In particular, if we define networks that are unusually difficult to control as those that are outliers in [Fig 4A and 4B](#) (shown as triangles, and defined as those with residuals $|\langle |CK| \rangle - \log_2 r| > 3\sigma$, with σ the standard deviation of those residuals across networks in each ensemble), [Eq 2](#) successfully corrects the predictions of mean control kernel sizes for those difficult cases.

The significance of [Eq 2](#) lies in the fact that, similar to [Eq 1](#), it enables an estimation of the average control kernel size solely based on the characteristics of the attractor landscape of the

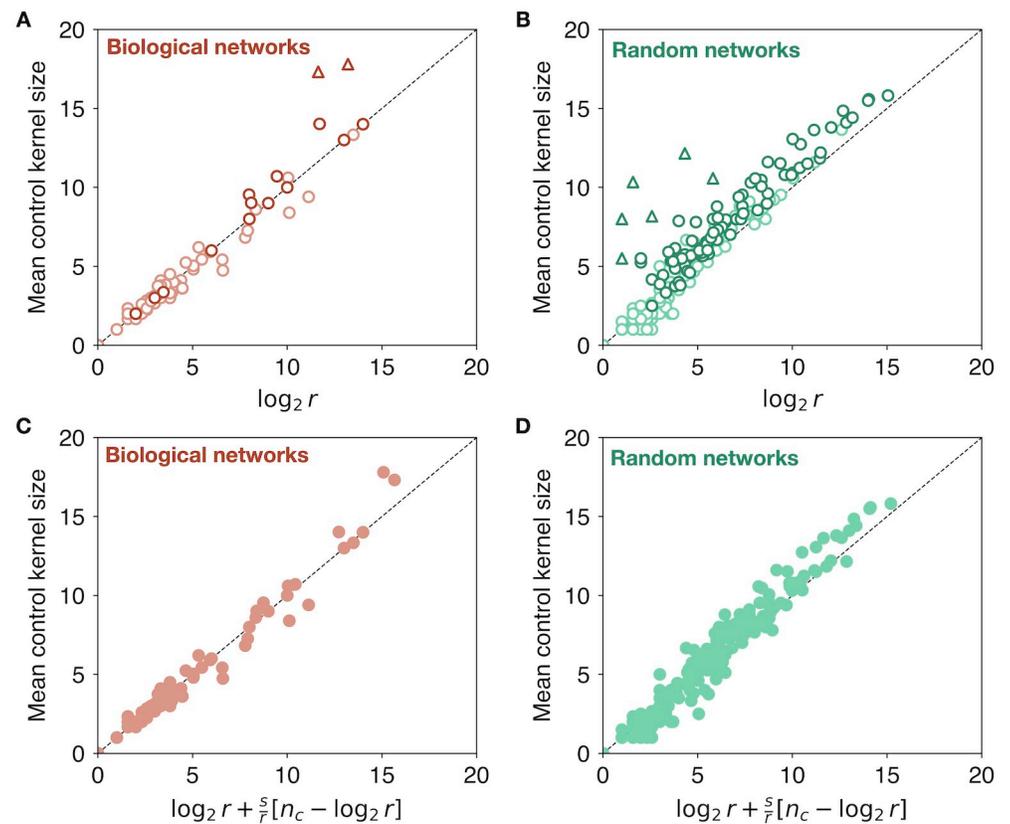


Fig 4. A correction that incorporates the number of isolated fixed points leads to better predictions of control kernel sizes. (A and B) The simple prediction that the mean control kernel size within each network is equal to the logarithm of the number of attractors r ([Eq 1](#)) works well for most networks, but there are a number of outlier networks that are much harder to control than expected. We mark networks as outliers (triangles) when their mean control kernel size is more than 3σ away from the expected value of $\log_2 r$, where σ is the standard deviation of the residuals $\langle |CK| \rangle - \log_2 r$. These outlier networks all have isolated fixed points (highlighted with a darker color). (C and D) A correction that assumes isolated fixed points have control kernel of size n_c ([Eq 2](#)) leads to better predictions. For the biological networks, the RMSE is 0.93 when using [Eq 1](#) and reduces to 0.60 when using [Eq 2](#). For the random networks, the RMSE reduces from 1.3 to 0.68.

<https://doi.org/10.1371/journal.pcsy.0000025.g004>

dynamics. This is achieved without the necessity of determining control kernels for all r attractors.

Fig 5 illustrates the initial variance in the distribution of $\langle |\text{CK}| \rangle$ across the analyzed biological and random networks. This variance can be better explained by our new predictor in Eq 2 as compared to the previous predictor in Eq 1. The improved precision in the new estimation is attained by integrating additional information, specifically the number s of isolated fixed points within the landscape, which is the primary cause for the deviations from our simple estimate.

Significantly, Eq 2 introduces a dependence of the control kernel on the network size (specifically on n_c). This seems to be at odds with our findings in Ref. [21], where we concluded that the average control kernel size remains independent of the network size for a given r . It is crucial to note that the parameter s is typically small in comparison to r , so that the influence of n_c is often hidden upon considering an average across all attractors in a network. This becomes especially apparent when Eq 2 is recast in the form

$$\langle |\text{CK}| \rangle \simeq \log_2 r + \frac{s}{r} \left[n_c - \log_2 r \right], \quad (3)$$

which shows how the size-dependent contribution is a correction to Eq 1 whose relevance scales with the fraction s/r of isolated fixed-points.

Once analyzed in the context of their biological relevance, we may expect the dependence on n_c to be even less impactful in real-world dynamics. The presence of isolated fixed-point attractors, which arise in models with deterministic updating rules, is undermined by the incorporation of stochasticity in more realistic models of gene regulation (see Discussion).

Finally, treating Eqs 1 and 2 as statistical predictors of average control kernel size, we examine the distributions of residuals away from these predictions. The non-normality of the residual distribution for the uncorrected Eq 1 is visible in Fig 5B, with the outlier networks contributing to a sizable skew. This gives a statistical motivation to look for a better predictor. Our corrected Eq 2 then produces a distribution of residuals that is closer to normal, with the outliers largely gone (Fig 5C). We quantify this approach to normality using the Shapiro-Wilk statistic: for the biological networks, W increases from 0.64 using Eq 1 to 0.85 using Eq 2, and for the random networks, W increases from 0.68 to 0.87. See S1 Text and S2 Fig for a more thorough discussion and visualization of the residual distributions.

3 Methods

Boolean Networks and attractor dynamics. In this section, we define key terms and concepts related to Boolean dynamical systems. Readers already familiar with Boolean network dynamics may skip to the next section.

The state of a Boolean network at time t is characterized by the states of its n nodes, $x_i(t)$, where $i = 1, \dots, n$. Due to the Boolean nature of the model, each node i can be in one of two possible states: ON, represented by $x_i(t) = 1$, or OFF, represented by $x_i(t) = 0$. In the context of cell regulation, the nodes typically correspond to interacting genes, with $x_i(t)$ serving as the Boolean approximation of the expression level of gene i . Here, the gene is considered expressed when $x_i(t) = 1$, and inactive otherwise.

The dynamics of a deterministic Boolean network are characterized by n Boolean functions f_1, \dots, f_n , which take as input the array $x_1(t), \dots, x_n(t)$ and produce the network's configuration

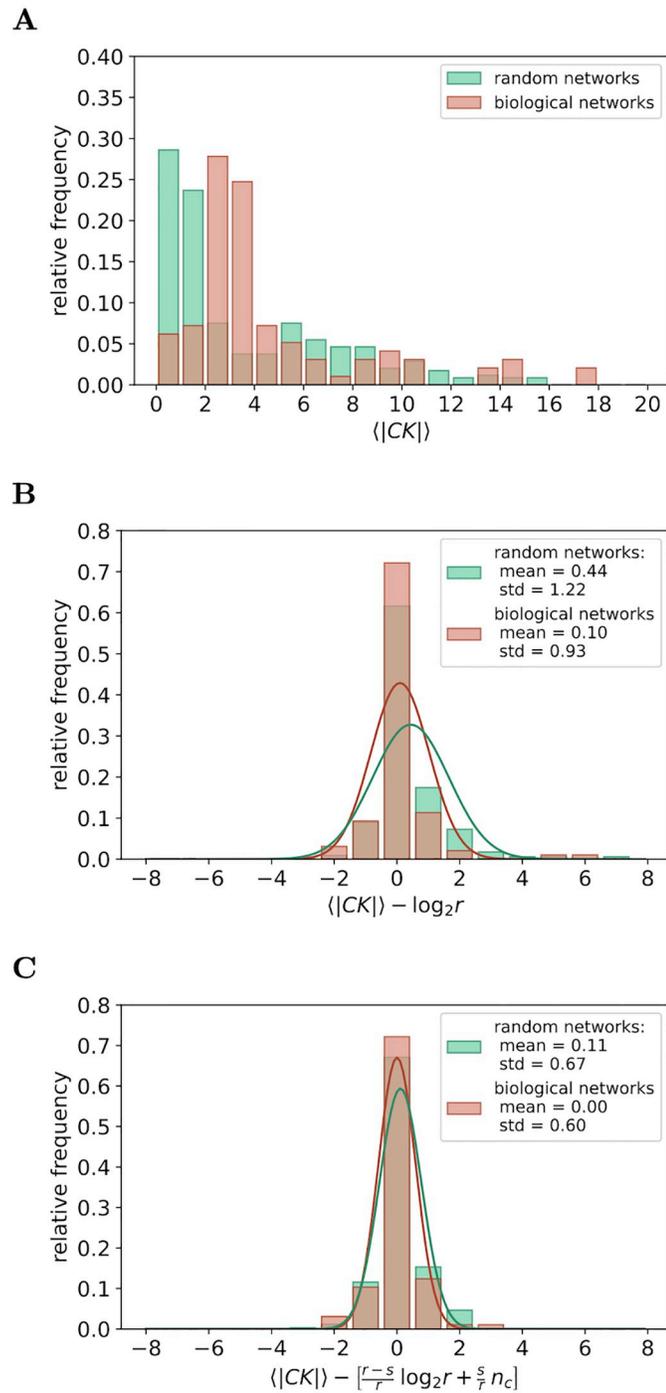


Fig 5. The isolated fixed point correction reduces the variance of mean control kernel size predictions. (A) The distributions of mean control kernel sizes across the ensembles of biological and random networks have large variances (listed in the legend). (B) Most of this variance can be explained in terms of the number of attractors r as in Eq 1. (C) A large fraction of the remaining variance is explained by using the corrected Eq 2 that incorporates the number of isolated fixed points s .

<https://doi.org/10.1371/journal.pcsy.0000025.g005>

at time $t + 1$:

$$\begin{aligned} x_1(t+1) &= f_1(x_1(t), \dots, x_n(t)) \\ &\dots \\ x_n(t+1) &= f_n(x_1(t), \dots, x_n(t)) . \end{aligned} \tag{4}$$

The specific form of these Boolean functions can be identified through experiments [27, 28]. This setup represents a *synchronous* update of all nodes in the network.

With n nodes, there are 2^n potential network configurations (i.e., gene expression patterns in the context of a genetic network). We will refer to these configurations as *states* of the network and the collection of all possible network states as the *configuration space*. The dynamics of the network are depicted by a time series of states. Although large, the configuration space is finite. Hence, when the functions f_i are deterministic, these dynamical trajectories will eventually settle into either a fixed state or a cycle of states. These specific sets of states are termed the *attractors* of the Boolean network. The collection of initial states that converge to a particular attractor is called its *basin of attraction*.

Control kernels. We characterize the difficulty of control using the notion of a control kernel [15]. Specifically, we define a control kernel for a given attractor as a set of nodes of minimal size that can be statically pinned to fixed values such that, in the pinned dynamics, all initial conditions converge to the desired attractor. Using this static definition of control, all fixed point attractors have a control kernel of some size, but not all cycles are controllable.

Many biological network models include nodes that represent external or environmental conditions. We call such nodes “input nodes,” and we implement their dynamics using the updating rule $x(t+1) = x(t)$. Input nodes are included in the network state, and we incorporate all possible values of these nodes when enumerating a network’s attractors. Based on our definition of control, all input nodes are included in every control kernel, as no other nodes can influence their state [21]. In rare cases, a network model will define a node’s dynamics such that it always results in a specific constant value C , that is, $x(t+1) = C$, with $C = 0$ or 1 . Such nodes have a fixed value across all possible attractors. They therefore do not distinguish different attractors and are thus never included as part of a control kernel.

For set environmental conditions (fixed states of input nodes) multiple attractors can only exist if the network structure presents feedback loops [29]. Any set of nodes connected in a tree structure without closed loops will depend deterministically on the states of input nodes in each attractor state. For this reason, nodes outside of a feedback loop will never be part of a control kernel (as all behavior of trees will be fully controlled once the attractor is selected within the core). For this reason, we remove from our analysis all dependent trees of nodes from each network, leaving only a number $n_c \leq n$ of “core” nodes that participate in feedback loops and have the potential to be part of control kernels. (A subtlety arises in computing n_c in that redundant expressions in the updating rules can create inoperative edges [30]. We detail in [S1 Text](#) why we do not expect this to affect our results.) Alternatively, we could include all n nodes in our analysis, but this would preclude the possibility of having any isolated fixed points in networks that have any nodes on which no other nodes depend (as changing the states of such nodes does not affect the dynamics).

Finally, we note that the core nodes of a network with no cycles consist only of the network’s input nodes. This corresponds to the fact that every attractor is fully specified by the states of the inputs, and results in the control kernel being equal to the set of input nodes. It is also possible for inputs to fully control networks that include cycles. In all of these cases, the number of attractors is 2^m , control kernels are of size m , and the logarithmic scaling of [Eq \(1\)](#)

is satisfied exactly. Therefore, these cases do not show up among those networks that we consider unusually difficult to control.

Network ensembles. For this study, we complement the set of network models analyzed in Ref. [21] with the more recent database compiled by Kadelka et al. [25] consisting of 122 Boolean models of biological regulation. The models in this database were chosen from a pool of 163 models identified using the Pubmed biomedical literature search engine, further restricted to include only expert-curated models (where both nodes and update rules were manually selected to prevent artifacts induced by prediction algorithms) and only one version of closely related models. Out of the 122 models, 61 were present in the Cell Collective database [31] that we analyzed in Ref. [21]. For consistency with our previous work, we retain a small number of Cell Collective networks that were not included in the Kadelka et al. database. The resulting set of networks encompasses models describing the regulatory logic governing various processes across a diverse array of species spanning different kingdoms of life, including animals, plants, fungi, and bacteria. For additional details, see Ref. [25] and its supplementary information.

A key limitation of Boolean dynamical models is the need to limit their size, as the number of network states grows exponentially with the number of nodes. To manage this, peripheral regions of the interactome are often simplified into input and output nodes, based on the assumption that the highlighted subsystems function relatively independently of the complete cellular interactome. While studies have shown that the network topology of these reduced models partially preserves the motif distribution of the full networks [32], unsurprisingly this preservation diminishes as the size of the extracted subgraphs decreases.

To increase the variance in our tested models, we also include three ensembles of random Boolean networks, as presented in Ref. [21]. These include networks with random truth tables, as well as Erdős-Rényi networks governed by threshold rules, with thresholds chosen either at zero total input or balanced to account for the number of inputs to each node. First, we use the well-studied ensemble of p - K random networks [33] to generate dynamics with random truth tables. Each node receives input from $K = 2$ nodes, with Boolean updating rules chosen such that the probability of the ON state is p . We sampled 225 networks from this ensemble, with 75 each having $p = 0.25, 0.5,$ and $0.75,$ and 25 each within these sets having number of nodes $n = 10, 15,$ and $20.$ We found control kernels for all controllable attractors in each of these sampled networks. Second, we define an ensemble that assigns each node a set of input nodes whose states are summed and compared to a threshold [15, 34]. The network's dependency structure A is chosen as an Erdős-Rényi graph with average degree $d,$ and each edge in this graph is assigned with probability p_I a value of -1 (representing an inhibitory interaction) or $+1$ (representing an excitatory interaction). Each node's state is determined by comparing the sum of the incoming signed inputs $s_i = \sum_j A_{ij}x_j(t)$ to its threshold τ_i :

$$x_i(t+1) = \begin{cases} 0, & \text{if } s_i < \tau_i \\ x_i(t), & \text{if } s_i = \tau_i \\ 1, & \text{if } s_i > \tau_i. \end{cases} \quad (5)$$

For one ensemble, we set thresholds to zero ($\tau_i = 0 \forall i$). In cases with little inhibition, these networks are biased toward excitation. A second ensemble consists of "balanced" threshold networks, with thresholds of individual nodes set at the center of the distribution of possible summed inputs ($\tau_i = \frac{1}{2} \sum_j A_{ij}$). We sampled 75 networks using each type of threshold, with each combination of $d = \{1, 2, 3, 4, 5\}, p_I = \{0.1, 0.3, 0.5, 0.7, 0.9\},$ and $n = \{10, 15, 20\},$ for a total of 150 networks. Of these, for 146 we were successful in finding control kernels for all controllable attractors.

The control results for all random networks included in this study are shown in [S1 Fig](#), where we replot [Fig 4B](#) to highlight the differences arising from the various methods used to construct the random networks. Notably, all outliers, marked as triangles in [Fig 4B](#), originate from the zero-threshold ensemble.

We use a slightly stricter criterion for choosing networks to include here than in previous work [[21](#)]. As before, we only include networks for which attractors and control kernels can be computed with a reasonable amount of computer time (about 1 day of compute time on a single CPU), and we further restrict ourselves to include only networks for which we are able to compute attractors exactly, without relying on sampling. We are able to compute exact attractors and control kernels for 104 biological networks (out of 122) and 371 random networks (out of 375). Furthermore, a limited number of networks (7 out of 104 biological networks and 25 out of 371 random networks) are characterized by uncontrollable cycles with basins that cumulatively comprise more than 99% of the state space. Due to the static nature of the control kernel examined in this study, we filter out these networks from further analysis.

Complete lists of networks analyzed in this work are included as comma separated files in the accompanying Zenodo data repository, and the biological networks are listed in [S1](#) and [S2](#) Tables.

Modularity. Biological regulatory network models often exhibit high modularity in their topology, where subsets of nodes interact more intensely within their own subsets than with other subsets. Additionally, these models tend to be hierarchical, featuring upstream modules that operate independently of downstream module behavior [[35](#)].

Efficient computation of attractors in modular and hierarchical Boolean networks can be achieved by analyzing modules separately. The approach involves identifying attractors for upstream modules first and subsequently leveraging each upstream attractor to determine corresponding attractors for downstream modules. Despite some intricacies discussed below, the computational execution of this process remains straightforward.

We first decompose a given Boolean dynamical system into hierarchical modules, defined as strongly connected components within the system's causal network. This representation is in the form of a directed acyclic graph, capturing dependencies among modules. We then identify attractors for upstream modules that are independent of others, utilizing a brute-force approach iterating over all possible states within the module.

For each downstream module \mathcal{D} , attractors \mathbf{U} of upstream modules, on which \mathcal{D} relies, are explored through all possible combinations. Given an attractor $U \in \mathbf{U}$, the states of nodes in upstream modules are fixed to the values present in U , and corresponding attractors for \mathcal{D} are determined using a brute-force approach iterating over all possible dynamics of \mathcal{D} based on the input from U .

Two subtleties arise when computing the set of all possible attractors \mathbf{U} for nodes upstream of \mathcal{D} . In the simplest case, where \mathcal{D} depends on separate modules $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ without shared ancestor modules, and all corresponding sets of upstream attractors $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$ consist only of fixed points, \mathbf{U} comprises $\prod_i |\mathbf{U}_i|$ attractors. However, when upstream modules share ancestor modules further upstream, or when dealing with cyclic attractors of length $\ell > 1$, additional considerations are necessary to account for inconsistent combinations and phase shifts between attractors, respectively.

We use this modular approach to efficiently identify attractors and compute control kernels. For more details, see Ref. [[21](#)].

Backward Reachable Sets. [Eq 2](#) relies on evaluating the basin sizes, a non-trivial task. To find basin sizes, one must solve for the equilibrium dynamics starting from every state of the system. The size of the network becomes a limiting factor due to the exponential growth of the number of states. While we can capture the entire set of attractor states for networks as large as

~ 300 nodes by leveraging the modular structure of their topology (see *Modularity* subsection above), this approach does not directly produce basin sizes. For smaller networks, we can directly compute them, but for the larger biological networks, this is infeasible. While we can sample a fixed number of initial conditions to estimate basin sizes, our inability to distinguish attractors that we never find in sampling from those with very small basin size limits the usefulness of this approach. Instead, we will leverage the knowledge of the attractors' identities and the fact that in Eq 2 we only need to know the number of them that are isolated.

In what follows, we present an efficient strategy for identifying the backwards reachable set of states in a Boolean network. We employ this method to quantify the exact number of isolated attractors s within the complete set of attractors r identified using the modular approach.

Let us refer to a generic Boolean network with n nodes and updating rules as in (4). We will call $S_{i,b}$ the subset of the configuration space S containing all the states satisfying the non-linear equation $f_i(x_1, \dots, x_n) = b$, where $b = 0, 1$. Therefore, given a state $X = (x_1, \dots, x_n)$, its backward reachable set, defined as the set of its pre-images, is

$$S_X = S_{1,x_1} \cap S_{2,x_2} \cap \dots \cap S_{n,x_n} . \tag{6}$$

Finding the $2n$ sets $S_{i,b}$ is equivalent to finding the pre-images of each state in S . A straightforward algorithm for finding $S_{i,b}$ would evaluate $f_i(x_1, \dots, x_n)$ for all $(x_1, \dots, x_n) \in S$, and select all the states for which $f_i(x_1, \dots, x_n) = b$. This approach would require $2n \times 2^n$ operations, as opposed to the $n \times 2^n$ operations we would need to determine the entire forward evolution of the network. In general, using a reverse algorithm is not a better way of solving the dynamics of a network [36]. Yet for our problem we only need to find the backward reachable sets for each known fixed point, to determine if its only pre-image is itself and it is therefore an isolated fixed point. From a computational standpoint, additional simplifications arise from the fact that the in-degree of the nodes in the network, i.e., the average number of inputs in the f_i functions, is often much smaller than n . Considering this, a more efficient method for performing the intersection in Eq 6 is to employ a parametric approach based on conditions solely related to the variables involved in each Boolean function.

If we call $P_{i,b}$ the set of sets of parametric equations defining the elements of $S_{i,b}$, then the set of sets of equations defining the elements of S_X is

$$P_X = P_{1,x_1} \bar{\times} P_{2,x_2} \bar{\times} \dots \bar{\times} P_{n,x_n} , \tag{7}$$

where $\bar{\times}$ denotes an *unordered* and *repetition-free* Cartesian product of sets, from which sets containing *inconsistent equations have been removed*.

Examples; Providing examples will help clarify our method. One benefit of referencing the parametric sets $P_{i,b}$ is that they serve as a highly concise representation of the sets $S_{i,b}$ when they encompass substantial subsets of the state space. For instance, an updating function with only k inputs will impose constraints solely on the k variables involved, leaving the remaining $n - k$ unconstrained. The number of combinations of these k variables will be multiplied by the 2^{n-k} number of combinations of the unconstrained variables. It is not uncommon in the biological networks we study to encounter functions with only one input out of tens of possible genes. In a scenario like this, having just twenty nodes can result in the order of the set $S_{i,b}$ being half a million, but its parametric set $P_{i,b}$ includes only one condition for its input.

Next, let us examine an example of how the products of the $P_{i,b}$ sets are calculated. For simplicity, we will refer to the smallest biological network in our database: a gene regulatory network model for mammalian cortical area development from Ref. [37]. This model comprises just five genes: *Coup_fti*, *Emx2*, *Fgf8*, *Sp8*, and *Pax6*. The network admits two fixed-point attractors, $X_1 = [0, 0, 1, 1, 1]$ and $X_2 = [1, 1, 0, 0, 0]$ (neither of them being isolated). Let us find

the pre-images of X_1 . From the updating rules

$$\begin{aligned} Coup_fti &= \overline{Sp8 \text{ OR } Fgf8} \\ Emx2 &= Coup_fti \text{ AND } \overline{Fgf8} \text{ AND } \overline{Pax6} \text{ AND } \overline{Sp8} \\ Fgf8 &= Fgf8 \text{ AND } Sp8 \text{ AND } \overline{Emx2} \\ Sp8 &= Fgf8 \text{ AND } \overline{Emx2} \\ Pax6 &= Sp8 \text{ AND } \overline{Coup_fti} \text{ AND } \overline{Emx2} \end{aligned}$$

(where an overline represents the Boolean operator NOT) we can deduce the parametric sets relevant to X_1 :

$$\begin{aligned} P_{Coup_fti,0} &= \{ \{ Sp8 = 1 \}, \\ &\quad \{ Fgf8 = 1, Sp8 = 0 \} \} \\ P_{Emx2,0} &= \{ \{ Coup_fti = 0 \}, \\ &\quad \{ Coup_fti = 1, Fgf8 = 1 \}, \\ &\quad \{ Coup_fti = 1, Fgf8 = 0, Pax6 = 1 \}, \\ &\quad \{ Coup_fti = 1, Fgf8 = 0, Sp8 = 1, Pax6 = 0 \} \} \\ P_{Fgf8,1} &= \{ \{ Fgf8 = 1, Emx2 = 0, Sp8 = 1 \} \} \\ P_{Sp8,1} &= \{ \{ Emx2 = 0, Fgf8 = 1 \} \} \\ P_{Pax6,1} &= \{ \{ Coup_fti = 0, Emx2 = 0, Sp8 = 1 \} \} . \end{aligned}$$

For example,

$$P_{Coup_fti,0} \times P_{Fgf8,1} = \{ \{ Fgf8 = 1, Emx2 = 0, Sp8 = 1 \} \} = P_{Fgf8,1} ,$$

because the products including the contradicting conditions $Sp8 = 0$ and $Sp8 = 1$ have been removed.

Once all five sets are multiplied, the result is

$$P_{X_1} = \{ \{ Coup_fti = 0, Emx2 = 0, Fgf8 = 1, Sp8 = 1 \} \} .$$

As no condition is set on $Pax6$,

$$S_{X_1} = \{ [0, 0, 1, 1, 0], [0, 0, 1, 1, 1] \} .$$

One of the two pre-images is just X_1 , as expected from that fact that it is a fixed-point.

For isolated fixed-points, this calculation returns a set of order one, where the only element is the fixed-point itself. Therefore, for each network, we perform this test on all of its fixed-points in order to count the number s of them that are isolated.

In computational terms, each set $P_{i,b}$ is essentially a dataframe. This dataframe has columns representing the input nodes of the function f_i , and each row corresponds to a combination of dynamic variables (node dynamical states) for which f_i evaluates to b_i . The solution to the problem involves obtaining the result sequentially by taking Cartesian products of all n dataframes and subsequently eliminating inconsistent rows.

A potential complication arises from a temporary combinatorial growth in the number of rows when products are computed between conditions over functions that do not share many inputs. This occurs when there are few cancellations. To address this, we initially conduct a Louvain community detection analysis on the network [38]. We then perform the first round of products among dataframes associated with functions that exhibit the highest overlap

among their inputs. Finally, we conduct the ultimate round of products among the results obtained in this manner. (This is computationally more efficient than using the modular structure for the first round of products, as it removes the limitation imposed by having a strict hierarchical structure.) This approach provides an efficient method for determining the backward reachable sets of our candidate isolated points.

4 Discussion

A comprehensive understanding of biological control is fundamental for both practical applications, such as genetic reprogramming [23, 39], regenerative medicine [40], and drug design [41–43], and for advancing our theoretical understanding of cellular coordination in systems biology [44]. As we explore the intricacies of controlling biological systems, essential questions arise: How easily can we exert control over these systems? What factors in a system’s design and behavior contribute to the ease or difficulty of control?

Alternative approaches to modeling the control of biological networks exist, and results can vary significantly depending on the goals set for the control. For example, the ability to attain any possible target state in a typical GRN described by a linear, continuous system requires control over more than 80% of the network [6]. Here we focus instead on the discrete, nonlinear case, and we use a definition of control that aims to force the system to a single one of its original attractor states from any initial condition.

With a few notable outliers, our previous work on this type of control indicated that the average required number of control nodes primarily depends on the number of preexisting attractors and shows no dependence on the system’s size. While these outliers were not numerous enough to undermine the general trend of ‘easy controllability’, they raised the question of whether biological systems exist that are inherently harder to control.

This work represents a more in-depth analysis of these outliers. Similarly to Ref. [21], we compare the results from our database of biological networks to those obtained in several ensembles of systems with randomly assigned network topologies and dynamic rules. These random ensembles are particularly important to include in that they produce a larger number of harder to control networks. Additionally, we roughly doubled the number of biological networks we tested (increasing from 49 to 104 models) through the inclusion of networks cataloged in Ref. [25]. Importantly, no new outliers from our original scaling law were observed.

Our primary observation here unveils a consistent pattern in these outlier networks, found in both the biological and random ensembles. This pattern originates from the presence of isolated fixed points—fixed points that attract no other states and prove to be among the attractors most challenging to control, with a control kernel that often includes all or almost all core nodes in the network.

Why is there a connection between unstable fixed points and difficult control? One simple way to get both instability and difficult control is exemplified in our random network ensemble that uses threshold updating rules. When thresholds are set uniformly low, this leads to a bias toward activation: the dynamics tend to push toward states with more 1s than 0s. States with few or no 1s can sometimes map to themselves, forming isolated fixed points. These states are also difficult to control because any node that is allowed to be active leads away toward further spreading of activation.

This type of bias is also seen in the few biological networks that are unusually difficult to control. Among the available biological networks, we found two outliers at the 3-sigma level (triangles in Fig 4A), both related to the ErbB network in breast cell lines: “SKBR3 Breast Cell Line Long-term ErbB Network” and “HCC1954 Breast Cell Line Long-term ErbB

Network.” These networks, which are part of the same study focused on resistance mechanisms in breast cancer treatments [45], all include isolated fixed points. Upon examining the attractors that are hardest to control (those with the maximum control kernel size), we find they consist of states where most nodes are inactive, except for a few nodes that self-sustain activity. This appears to be due to a highly biased network core where the all-zero state is an isolated fixed point, coupled with peripheral nodes that can independently sustain themselves. This combination increases the number of hard-to-control fixed points and thereby the average control kernel size. It is worth noting that this ability of peripheral nodes to sustain their own activity may be an effect of the algorithm used in the original study to infer the network logic.

Unfortunately, this connection between unstable fixed points and difficult control itself has exceptions. In particular, it is possible to construct networks that are difficult to control without having globally unstable (isolated) fixed points. (We describe one such counterexample in [S1 Text](#), also illustrated in the last row of [Fig 1](#), a network dynamic in which a locally unstable fixed point has maximally distant network states that map back to it. We find only one similar network in the random ensembles that we study, and that network also contains a globally unstable fixed point. Unsurprisingly, we do not find any network resembling such a contrived counterexample among the biological systems.) Still, by efficiently counting the unstable fixed points of a network, we are able to better explain the variance in control kernel sizes by incorporating the prevalence of these unstable fixed points. This constitutes a useful tool to predict the expected amount of required control over a network based solely on partial information of the attractor landscape. While determining the number of attractors in a Boolean network is itself a computationally hard problem (#P-complete [46]), once this is known, our approach avoids the NP-hard problem of evaluating individual control kernels.

The observation that difficult control is related to unstable fixed points is reassuring for several reasons. Firstly, these fixed points are not a general feature of models with non-deterministic updating rules. Most sources of stochasticity [47] would push the dynamics toward more stable attractors characterized by larger basins. A weak form of stochasticity arises in asynchronous updating. In this case the identity of fixed-point attractors remains unchanged, but basin sizes can change. As a result, it is possible that isolated fixed points would no longer be isolated. A stronger form of stochasticity would include bit-flips. In this case, we expect that isolated fixed points would effectively disappear. More importantly, the incompatibility of unstable fixed points with stochasticity makes it highly improbable for such states to carry biological significance, i.e., to represent actual cell types. Their persistence would require fine-tuned preservation of an unstable dynamical state.

The biological irrelevance of such unstable fixed points is also significant for another reason. As they reintroduce an explicit dependence on the network size, these cases of difficult control would contradict our main statement that control does not scale with the size of the system [21], in direct disagreement with recent results in the genetic reprogramming of mammalian cells [23]. In this light, we interpret our correction ([Eq 2](#)) as an empirical method for obtaining a more precise estimate of the required control when employing a simplistic, deterministic model of the regulatory dynamics. Verifying that our size-independent scaling law remains valid in models of genetic regulation that include stochasticity is needed but beyond the scope of this work.

More generally, given our results one might hope that the amount of control necessary to select macroscopic phenotypes could be approximated without knowing details of the microscopic dynamics.

Supporting information

S1 Fig. Mean control kernel sizes as a function of the logarithm of the number of attractors for three random network ensembles. Here we replot Fig 4B to emphasize differences that arise due to constructing random networks in different ways: using random truth tables (dark triangles), threshold networks with balanced threshold (light squares), and threshold networks with zero threshold (green circles). Note that all the outliers identified as triangles in Fig 4B come from the zero threshold ensemble. This figure omits 4 networks that were included in Fig 9A of Ref. [21] due to our more strict threshold that excludes networks that have uncontrollable cycles with large basins.

(PDF)

S2 Fig. Quantile–quantile (Q–Q) plots of the residual distributions shown in Fig 5B and 5C. In all four cases displayed, the vertical axes (*ordered quantiles*) represent the observed residuals, and the horizontal axes (*theoretical quantiles*) represent the expected values assuming normally distributed residuals. A perfect match to normality would align with the red diagonal line, indicating linear scaling. The two plots on the left correspond to predictions using logarithmic scaling, while the plots on the right use our corrected scaling. The top row shows results for biological networks, and the bottom row shows results for random networks.

(PDF)

S3 Fig. Histograms of control kernel size for individual random networks with at least one isolated fixed point. Control kernel sizes of all attractors are displayed with unfilled bars, and the subset corresponding to isolated fixed points are shown as shaded bars. A star is shown at n_c for those networks that have at least one control kernel of size n_c (73 networks), and an X is shown at n_c otherwise (14 networks).

(PDF)

S4 Fig. Histograms of control kernel size for individual biological networks with at least one isolated fixed point. Control kernel sizes of all attractors are displayed with unfilled bars, and the subset corresponding to isolated fixed points are shown as shaded bars. Networks without loops have histograms depicted with hatching, as in Fig 3. A star is shown at n_c for those networks that have at least one control kernel of size n_c (16 networks), and an X is shown at n_c otherwise (1 network).

(PDF)

S1 Table. Cell Collective networks [31].

(PDF)

S2 Table. Additional biological networks from [25].

(PDF)

S1 Text. Supporting text. This includes details regarding the construction of a particular exceptional network, redundant regulators, and the shapes of residual distributions.

(PDF)

Acknowledgments

The authors acknowledge Research Computing at Arizona State University for providing HPC resources that have contributed to the research results reported in this paper. We thank Zhao Yuanchen for correcting errors in a previous version of Fig 1.

Author Contributions

Conceptualization: Bryan C. Daniels, Enrico Borriello.

Data curation: Bryan C. Daniels, Enrico Borriello.

Formal analysis: Bryan C. Daniels, Enrico Borriello.

Investigation: Bryan C. Daniels, Enrico Borriello.

Methodology: Bryan C. Daniels, Enrico Borriello.

Project administration: Bryan C. Daniels, Enrico Borriello.

Resources: Bryan C. Daniels, Enrico Borriello.

Software: Bryan C. Daniels, Enrico Borriello.

Supervision: Bryan C. Daniels, Enrico Borriello.

Validation: Bryan C. Daniels, Enrico Borriello.

Visualization: Bryan C. Daniels, Enrico Borriello.

Writing – original draft: Bryan C. Daniels, Enrico Borriello.

Writing – review & editing: Bryan C. Daniels, Enrico Borriello.

References

1. Kitano H. Computational systems biology. *Nature*. 2002; 420(6912):206–10. <https://doi.org/10.1038/nature01254> PMID: 12432404
2. Richardson SS, Stevens H. Postgenomics: Perspectives on biology after the genome. Duke University Press; 2015.
3. Isidori A. Nonlinear control systems. Springer Science & Business Media; 2013.
4. Borriello E, Walker SI, Laubichler MD. Cell phenotypes as macrostates of the GRN dynamics. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. 2020; 334(4):213–24. <https://doi.org/10.1002/jez.b.22938> PMID: 32157818
5. Zañudo JGT, Yang G, Albert R, Levine H. Structure-based control of complex networks with nonlinear dynamics. *Proceedings of the National Academy of Sciences of the United States of America*. 2017; 114(28):7234–9. <https://doi.org/10.1073/pnas.1617387114> PMID: 28655847
6. Liu YY, Slotine JJ, Barabási AL. Controllability of complex networks. *nature*. 2011; 473(7346):167–73. <https://doi.org/10.1038/nature10011> PMID: 21562557
7. Correia RB, Gates AJ, Wang X, Rocha LM. CANA: A Python Package for Quantifying Control and Canalization in Boolean Networks. *Frontiers in Physiology*. 2018 Aug; 9:1046. Available from: <https://www.frontiersin.org/article/10.3389/fphys.2018.01046/full>. PMID: 30154728
8. El-Samad H, Khammash M. Modelling and analysis of gene regulatory network using feedback control theory. *International Journal of Systems Science*. 2010 Jan; 41(1):17–33. Available from: <http://www.tandfonline.com/doi/abs/10.1080/00207720903144545>.
9. Biane C, Delaplace F. Causal Reasoning on Boolean Control Networks Based on Abduction: Theory and Application to Cancer Drug Discovery. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2019 Sep; 16(5):1574–85. Available from: <https://ieeexplore.ieee.org/document/8585043/>. PMID: 30582550
10. Beneš N, Brim L, Huvar O, Pastva S, Šafránek D, Šmířáková E. AEON.py: Python library for attractor analysis in asynchronous Boolean networks. *Bioinformatics*. 2022 Oct; 38(21):4978–80. Available from: <https://academic.oup.com/bioinformatics/article/38/21/4978/6697883>. PMID: 36102786
11. Su C, Pang J. CABEAN: a software for the control of asynchronous Boolean networks. *Bioinformatics*. 2021 May; 37(6):879–81. Available from: <https://academic.oup.com/bioinformatics/article/37/6/879/5897411>. PMID: 32845335
12. Videla S, Saez-Rodriguez J, Guziolowski C, Siegel A. caspo: a toolbox for automated reasoning on the response of logical signaling networks families. *Bioinformatics*. 2017 Mar; 33(6):947–50. Available from: <https://academic.oup.com/bioinformatics/article/33/6/947/2585024>. PMID: 28065903

13. Rozum JC, Deritei D, Park KH, Tejeda JG, Albert R. pystablemotifs: Python library for attractor identification and control in Boolean networks. *Bioinformatics*. 2021.
14. Murrugarra D, Veliz-Cuba A, Aguilar B, Arat S, Laubenbacher R. Modeling stochasticity and variability in gene regulatory networks. *EURASIP Journal on Bioinformatics and Systems Biology*. 2012 Dec; 2012(1):5. Available from: <https://bsb-eurasipjournals.springeropen.com/articles/10.1186/1687-4153-2012-5>. PMID: 22673395
15. Kim J, Park SM, Cho KH. Discovery of a kernel for controlling biomolecular regulatory networks. *Scientific reports*. 2013; 3:2223. <https://doi.org/10.1038/srep02223> PMID: 23860463
16. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*. 1969; 22(3):437–67. [https://doi.org/10.1016/0022-5193\(69\)90015-0](https://doi.org/10.1016/0022-5193(69)90015-0) PMID: 5803332
17. Zañudo JGT, Albert R. Cell Fate Reprogramming by Control of Intracellular Network Dynamics. *PLoS Computational Biology*. 2015; 11(4):1–24. <https://doi.org/10.1371/journal.pcbi.1004193> PMID: 25849586
18. Fiedler B, Mochizuki A, Kurosawa G, Saito D. Dynamics and Control at Feedback Vertex Sets. I: Informative and Determining Nodes in Regulatory Networks. *Journal of Dynamics and Differential Equations*. 2013; 25(3):563–604. <https://doi.org/10.1007/s10884-013-9312-7>
19. Mochizuki A, Fiedler B, Kurosawa G, Saito D. Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. *Journal of Theoretical Biology*. 2013; 335:130–46. Available from: <http://dx.doi.org/10.1016/j.jtbi.2013.06.009>. PMID: 23774067
20. Akutsu T, Hayashida M, Ching WK, Ng MK. Control of Boolean networks: Hardness results and algorithms for tree structured networks. *Journal of theoretical biology*. 2007; 244(4):670–9. <https://doi.org/10.1016/j.jtbi.2006.09.023> PMID: 17069859
21. Borriello E, Daniels BC. The basis of easy controllability in Boolean networks. *Nature communications*. 2021; 12(1):5227. <https://doi.org/10.1038/s41467-021-25533-3> PMID: 34471107
22. Kushilevitz E, Linial N, Rabinovich Y, Saks M. Witness sets for families of binary vectors. *Journal of Combinatorial Theory, Series A*. 1996; 73(2):376–80. [https://doi.org/10.1016/S0097-3165\(96\)80015-X](https://doi.org/10.1016/S0097-3165(96)80015-X)
23. Müller FJ, Schuppert A. Few inputs can reprogram biological networks. *Nature*. 2011; 478(7369):2–3. PMID: 22012402
24. Hou W, Ruan P, Ching WK, Akutsu T. On the number of driver nodes for controlling a Boolean network when the targets are restricted to attractors. *Journal of theoretical biology*. 2019; 463:1–11. <https://doi.org/10.1016/j.jtbi.2018.12.012> PMID: 30543810
25. Kadelka C, Butrie TM, Hilton E, Kinseth J, Schmidt A, Serdarevic H. A meta-analysis of Boolean network models reveals design principles of gene regulatory networks. *arXiv preprint arXiv:200901216*. 2020.
26. Klemm K, Bornholdt S. Stable and unstable attractors in Boolean networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*. 2005; 72(5):1–4. PMID: 16383673
27. Henry A, Monéger F, Samal A, Martin OC. Network function shapes network structure: the case of the arabidopsis flower organ specification genetic network. *Molecular BioSystems*. 2013; 9(7):1726–35. <https://doi.org/10.1039/c3mb25562j> PMID: 23579205
28. Zhou JX, Samal A, d'Hérouël AF, Price ND, Huang S. Relative stability of network states in Boolean network models of gene regulation in development. *Biosystems*. 2016; 142:15–24. <https://doi.org/10.1016/j.biosystems.2016.03.002> PMID: 26965665
29. Zañudo JGT, Yang G, Albert R. Structure-based control of complex networks with nonlinear dynamics. *Proceedings of the National Academy of Sciences*. 2017; 114(28):7234–9. <https://doi.org/10.1073/pnas.1617387114> PMID: 28655847
30. Gates AJ, Correia RB, Wang X, Rocha LM. The effective graph reveals redundancy, canalization, and control pathways in biochemical regulation and signaling. *Proceedings of the National Academy of Sciences of the United States of America*. 2021; 118(12). <https://doi.org/10.1073/pnas.2022598118> PMID: 33737396
31. Helikar T, Kowal B, McClenathan S, Bruckner M, Rowley T, Madrahimov A, et al. The Cell Collective: toward an open and collaborative approach to systems biology. *BMC systems biology*. 2012; 6:96. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3443426&tool=pmcentrez&rendertype=abstract>. PMID: 22871178
32. Borriello E. The local topology of dynamical network models for biology. *Journal of Complex Networks*. 2024; 12(2):cnae007. <https://doi.org/10.1093/comnet/cnae007>
33. Shmulevich I, Kauffman SA. Activities and sensitivities in Boolean network models. *Physical Review Letters*. 2004; 93(4):048701–1. <https://doi.org/10.1103/PhysRevLett.93.048701> PMID: 15323803

34. Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(14):4781–6. <https://doi.org/10.1073/pnas.0305937101> PMID: 15037758
35. Paul S, Su C, Pang J, Mizera A. A decomposition-based approach towards the control of Boolean networks. In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*; 2018. p. 11–20.
36. Wuensche A, Lesser M, Lesser MJ. *Global dynamics of cellular automata: an atlas of basin of attraction fields of one-dimensional cellular automata*. vol. 1. Andrew Wuensche; 1992.
37. Giacomantonio CE, Goodhill GJ. A Boolean model of the gene regulatory network underlying Mammalian cortical area development. *PLoS computational biology*. 2010; 6(9):e1000936. <https://doi.org/10.1371/journal.pcbi.1000936> PMID: 20862356
38. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*. 2008; 2008(10):P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
39. Kamimoto K, Adil MT, Jindal K, Hoffmann CM, Kong W, Yang X, et al. Gene regulatory network reconfiguration in direct lineage reprogramming. *Stem Cell Reports*. 2023; 18(1):97–112. <https://doi.org/10.1016/j.stemcr.2022.11.010> PMID: 36584685
40. Tewary M, Shakiba N, Zandstra PW. Stem cell bioengineering: building from stem cell biology. *Nature Reviews Genetics*. 2018; 19(10):595–614. <https://doi.org/10.1038/s41576-018-0040-z> PMID: 30089805
41. Ghosh S, Basu A. Network medicine in drug design: implications for neuroinflammation. *Drug discovery today*. 2012; 17(11–12):600–7. <https://doi.org/10.1016/j.drudis.2012.01.018> PMID: 22326234
42. Fortney K, Xie W, Kotlyar M, Griesman J, Kotseruba Y, Jurisica I. NetwORx: connecting drugs to networks and phenotypes in *Saccharomyces cerevisiae*. *Nucleic acids research*. 2012; 41(D1):D720–7. <https://doi.org/10.1093/nar/gks1106> PMID: 23203867
43. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *science*. 2006; 313(5795):1929–35. <https://doi.org/10.1126/science.1132939> PMID: 17008526
44. Davidson EH. *The regulatory genome: gene regulatory networks in development and evolution*. Elsevier; 2010.
45. Der Heyde SV, Bender C, Henjes F, Sonntag J, Korf U, Beißbarth T. Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines. *BMC Systems Biology*. 2014 Dec; 8(1):75. Available from: <https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-8-75>.
46. Kosub S. Dichotomy Results for Fixed-Point Existence Problems for Boolean Dynamical Systems. *Mathematics in Computer Science*. 2008 Mar; 1(3):487–505. Available from: <http://link.springer.com/10.1007/s11786-007-0038-y>.
47. Shmulevich I, Aitchison JD. Deterministic and stochastic models of genetic regulatory networks. *Methods in enzymology*. 2009; 467:335–56. [https://doi.org/10.1016/S0076-6879\(09\)67013-0](https://doi.org/10.1016/S0076-6879(09)67013-0) PMID: 19897099