# Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints

Victor Chardès, Massimo Vergassola, Aleksandra M. Walczak,* and Thierry Mora*
*Laboratoire de physique de l'École normale supérieure, CNRS, PSL University, Sorbonne Université, and Université de Paris, 75005 Paris, France*

In order to target threatening pathogens, the adaptive immune system performs a continuous reorganization of its lymphocyte repertoire. Following an immune challenge, the B cell repertoire can evolve cells of increased specificity for the encountered strain. This process of affinity maturation generates a memory pool whose diversity and size remain difficult to predict. We assume that the immune system follows a strategy that maximizes the long-term immune coverage and minimizes the short-term metabolic costs associated with affinity maturation. This strategy is defined as an optimal decision process on a finite dimensional phenotypic space, where a pre-existing population of naive cells is sequentially challenged with a neutrally evolving strain. We unveil a trade-off between immune protection against future strains and the necessary reorganization of the repertoire. This plasticity of the repertoire drives the emergence of distinct regimes for the size and diversity of the memory pool, depending on the density of naive cells and on the mutation rate of the strain. The model predicts power-law distributions of clonotype sizes observed in data, and rationalizes antigenic imprinting as a strategy to minimize metabolic costs while keeping good immune protection against future strains.

## I. INTRODUCTION

Adaptive immunity relies on populations of lymphocytes expressing diverse antigen-binding receptors on their surface to defend the organism against a wide variety of pathogens. B lymphocytes rely on a two-step process to produce diversity: first a diverse naive pool of cells is generated; upon recognition of a pathogen the process of affinity maturation allows B cells to adapt their B-cell receptor (BCR) to epitopes of the pathogen through somatic hypermutation [1]. This process, which takes place in germinal centers [2], can increase the affinity of naive BCR for the target antigen by up to a thousand fold factor [3]. Through affinity maturation, the immune system generates high-affinity, long-lived plasma cells, providing the organism with humoral immunity to pathogens through the secretion of antibodies—the soluble version of the matured BCR—as well as a pool of memory cells with varying affinity to the antigens [4]. However, the diversity and coverage of the memory pool, as well as the biological constraints that control its generation, have not yet been fully explored.

Analysis of high-throughput BCR sequencing data has revealed long tails in the distribution of clonotype abundances, identifying some very abundant clonotypes as well as many very rare ones [5, 6]. Additionally, many receptors have similar sequences and cluster into phylogenetically related lineages [7–11]. These lineages have been used to locally trace the evolution of antibodies in HIV patients [12, 13] and in influenza vaccinees [14, 15]. Analysis of influenza booster vaccines shows large clonal dominance, with 90% receptor sequences belonging to 10% of clones [16]. Memory B-cell clones are more diverse and less specific to the infecting antigen than antibody-producing plasma cells [17, 18]. This suggests that the immune system is trying to anticipate infections by related pathogens or future escape mutants [19].

Theoretical approaches have attempted to qualitatively describe affinity maturation as a Darwinian co-evolutionary process, and studied optimal affinity maturation schemes [20, 21], as well as optimal immunization schedules to stimulate antibodies with large neutralizing capabilities [22–24]. Most of these approaches have been limited to short timescales, often with the goal of understanding the evolution of broadly neutralizing antibodies. Here we propose a mathematical framework to explore the trade-offs that control how the large diversity of memory cells evolves over a lifetime.

Despite long-lasting efforts to describe the co-evolution of pathogens and hosts immune systems [25–29], and recent theoretical work on optimal schemes for using and storing memory in the presence of evolving pathogens [30], few theoretical works have described how the B-cell memory repertoire is modified by successive immunization challenges. Early observations in humans [31] have shown that sequential exposure to antigenically drifted influenza strains was more likely to induce an immune response strongly directed towards the first strain the patients were exposed to [32]. This immune imprinting with viral strains encountered early in life was initially called "original antigenic sin," as it can limit the efficiency of vaccination [33]. This phenomenon has been observed in a variety of animal models and viral strains [34]. Secondary infections with an antigenically diverged influenza strain can reactivate or "backboost" memory cells specific to the primary infecting strain [35]. This response is characterized by lower binding affinity but can still have in-vivo efficiency thanks of cross-reactive antibodies [36]. There is a long-standing debate about how detrimental "original antigenic sin" is [37, 38]. However, the question

*Corresponding authors. These authors contributed equally.

of under what circumstances an immune response based on memory re-use is favourable has not been addressed.

We build a theoretical framework of joint virus and repertoire evolution in antigenic space, and investigate how infections by evolving pathogens have shaped, over evolutionary timescales, the B-cell repertoire response and re-organization. Pathogens causing acute infections usually show a seasonal periodicity, while the maturation processes in the B-cell repertoire take place over a few weeks. This observation allows us to consider that affinity maturation happens in a negligible time with respect to the reinfection period. Within this approximation, we investigate the optimal immune maturation strategies using a framework of discrete-time decision process. We show the emergence of three regimes—monoclonal memory response, polyclonal memory response, and naive response—as trade-offs between immune coverage and resource constraint. Additionally, we demonstrate that reactivation of already existing memory clonotypes can lead to self-trapping of the immune repertoire to low reactivity clones, opening the way for "original antigenic sin."

## II. RESULTS

### A. Affinity maturation strategies for recurring infections

B cells recognize pathogens through the binding of their BCR to parts of the pathogen's proteins, called antigens. To model this complex protein-protein interaction problem, we assume that both receptors and antigens may be projected into an effective, $d$-dimensional antigenic space (Fig. 1), following the "generalized shape space" idea pioneered by Perelson and Oster [39]. Receptor-antigen pairs at close distance in that space bind well, while those that are far away bind poorly. Specifically, we define a cross-reactivity function $0 \leq f \leq 1$ quantifying the binding affinity between antigen $a$ and receptor $x$, which we model by a stretched exponential:

$$f(x, a) = e^{-(\|x-a\|/r_0)^q}, \tag{1}$$

where $r_0$ is the cross-reactivity radius, and $q$ a stretching exponent.

For simplicity, we focus on a single pathogen represented by its immunodominant antigen, so that each viral strain is represented by a single point $a_t$ in antigenic space (black square), where $t = 1, 2, \ldots$ is a discrete time counting the number of re-infections. It is difficult to estimate the rate of re-infections or exposures to the same pathogen. It can be fairly high in humans, where individuals are exposed to the most common viruses from less than once to several times a year [40]. The numbers of lifetime exposures would then range from a few to a few hundreds.

The B cell repertoire, on the other hand, is represented by a collection of antigenic coordinates corresponding to

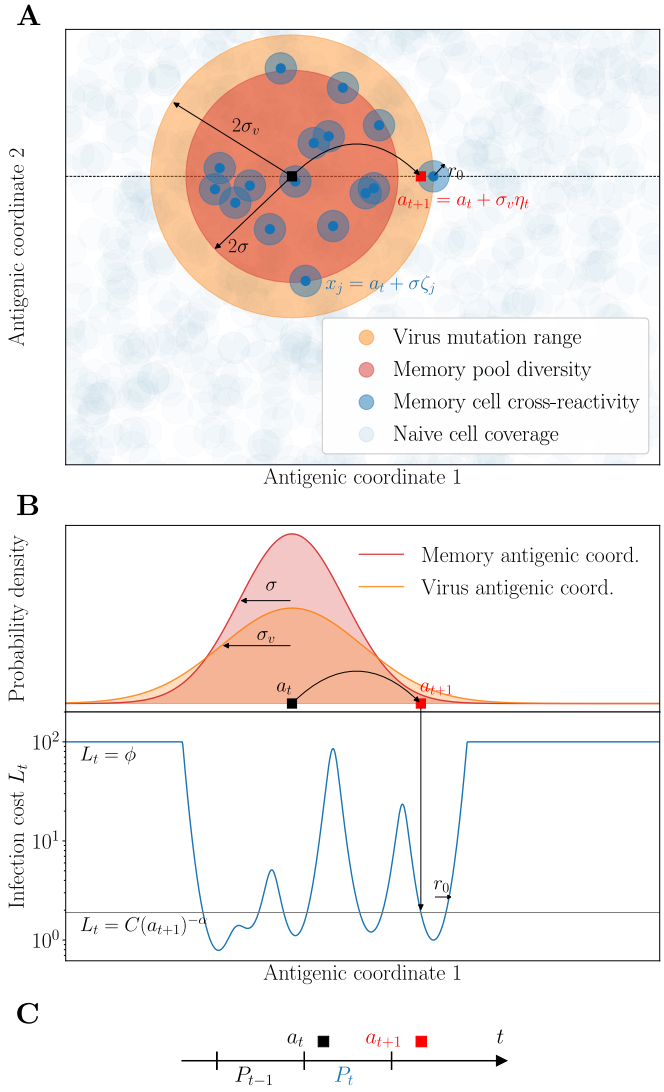


FIG. 1: **Model of sequential affinity maturation. A.** An infecting strain is defined by its position $a_n$ in antigenic space (dark square). In response, the immune system creates $m$ new memory clonotypes $x_j$ (blue points) from a Gaussian distribution of width $\sigma$ centered in $a_t$ (red area). These new clonotypes create a cost landscape (blue areas) for the next infection, complemented by a uniform background of naive cells (light blue). The next infecting strain (red square) is drawn from a Gaussian distribution of width $\sigma_v$ centered in $a_t$ (orange area). The position of this strain on the infection landscape is shown with the arrow. Antigenic space is shown in 2 dimensions for illustration purposes, but can have more dimensions in the model. **B.** Cross-section of the distributions of memories and of the next strain, along with the infection cost landscape $L_t$ (in blue). Memories create valleys in the landscape, on a background of baseline naive protection $\phi$. **C.** Sequential immunization. Strain $a_t$ modifies the memory repertoire into $P_t$, which is used to fight the next infection $a_{t+1}$. $P_t$ is made of all newly created clonotypes (blue points in A) as well as some previously existing ones (not shown). Clonotype abundances are boosted following each infection as a function of the cross-reactivity, and each individual cell survives from one challenge to the other with a probability $\gamma$.

each receptor clonotype. We distinguish memory cells (dark blue circles in Fig. 1A), denoted by $P_t$, which have emerged in response to the presence of the virus, and a dense background of naive cells $N$ (light blue circles) which together provide a uniform but weakly protective coverage of any viral strain (subsumed into the parameter $\phi$ defined later).

The viral strain evolves randomly in antigenic space, sequentially challenging the existing immune repertoire. This assumption is justified by the fact that for acute infections with a drifting viral strain, such as influenza, the immune pressure exerted on the strain does not happen in hosts but rather at the population level [25]. Viral evolution is not neutral, but it is unpredictable from the point of view of individual immune systems. Specifically, we assume that, upon reinfection, the virus is represented by a new strain, which has moved from the previous antigenic position $a_t$ to the new one $a_{t+1}$ according to a Gaussian distribution with typical jump size $\sigma_v$, called "divergence":

$$a_{t+1} = a_t + \sigma_v \eta_{t+1}, \tag{2}$$

where $\langle \eta_t \rangle = 0$ and $\langle \eta_t \cdot \eta_{t'} \rangle = \delta_{tt'}$.

To model how the immune repertoire is re-organized in response to a new infection $a_t$, we postulate that its strategy has been adapted over evolutionary timescales to maximize the speed of immune response to subsequent challenges, given the resource constraints imposed by affinity maturation [41]. This strategy is assumed to be stochastic and characterized by the law $\mathcal{S}$ according to which the BCR repertoire evolves from $P_{t-1}$ to $P_t$ as a result of affinity maturation: $P_t \sim \mathcal{S}(P_{t-1}, a_t)$ (Fig. 1C).

We consider the following rules inspired by known mechanisms available to the immune system [2]. Upon infection by $a_t$, $m_t$ new receptors are matured to target the virus, with $m_t$ distributed according to a Poisson law of mean $\bar{m}$. Each new receptor is roughly located around $a_t$ in antigenic space with some imprecision $\sigma$: $x_j = a_t + \sigma \xi_j$, $j = 1, \ldots, m_t$, with $\xi_j$ normally distributed with $\langle \xi_j \rangle = 0$ and $\langle \xi_j^2 \rangle = 1$. These new positions $(x_j)_{j=1}^{m_t}$ are added to the ones from the previous repertoire $P_{t-1}$.

In addition, each clonotype $x \in P_{t-1}$ from the previous repertoire may be reactivated and be subsequently duplicated through cell divisions [19], with probability $\mu f(x, a_t)$, proportional to the cross-reactivity given by (1), where $0 \leq \mu \leq 1$ is a proliferation parameter. These previously existing cells and their offspring may then die before the next infection. We denote by $\gamma$ their survival probability, so that the average lifetime of each cell is $(1-\gamma)^{-1}$. We denote by $c_{x,t}$ the size of clonotype $x$ at time $t$. By convention we set $c_{x,t} = 1$ for newly matured clonotyes.

This strategy $\mathcal{S}$ is characterized by two parameters that can be tuned by evolution: the mean number $\bar{m}$ and diversity $\sigma$ of newly matured clonotypes. The proliferation and death parameters $\mu$ and $\gamma$ are assumed to be constrained and fixed. To assess the performance of a given strategy, we define an overall loss at each time step:

$$L_t = I_t + \kappa m_t = \min \Big[ \phi, \Big( \sum_{x \in P_{t-1}} c_{x,t} f(x, a_t) \Big)^{-\alpha} \Big] + \kappa m_t, \tag{3}$$

where $\phi > 0$ is a maximal cost corresponding to using naive cells, and $\alpha > 0$ and $\kappa > 0$ are model parameters.

The first term of (3) corresponds to the infection cost $I_t$, which is a decreasing function of the coverage of the virus by the memory repertoire, $C(a_t) = \sum_{x \in P_{t-1}} c_{x,t} f(x, a_t)$, with a power $\alpha$ governing how sharp that dependence is. Intuitively, the lower the coverage, the longer it will take for memory cells to mount an efficient immune response and clear the virus, incurring larger harm on the organism [41, 42]. When memory coverage is too low, the naive repertoire takes over, incurring a maximal cost fixed to $\phi$. In the Supplementary Text we show how this naive cut-off may be derived in a model where the immune system can choose to activate either its memory or its naive compartment in response to a new infection, when naive clonotypes are very numerous but offer weak protection. In that interpretation, $\phi$ scales like the inverse density of naive cells, so that we will refer to $\phi^{-1}$ as "naive density." In Fig. 1B we plot an example of the infection cost along a cross-section of the antigenic space.

The second term in the right-hand side of (3) corresponds to a plasticity cost encoding the resources necessary to implement affinity maturation. It is assumed to be proportional to the number of new clonotypes $m_t$, with cost factor $\kappa$.

We assume that, over evolutionary timescales, the immune system has minimized the average cumulative cost over a large number of infections $T$:

$$\mathcal{L}(\mathcal{S}) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} L_t. \tag{4}$$

This optimization yields the optimal parameters of the strategy. In the case described above, we obtain the optimal mean number of added receptors $\bar{m}^*$ and their optimal diversity $\sigma^*$:

$$(\bar{m}^*, \sigma^*) = \arg\min_{(m,\sigma)} \mathcal{L}(\mathcal{S}). \tag{5}$$

In the next section we discuss the results of that optimization with various choices of parameters.

### B. Phase diagram of optimal affinity maturation strategies

We obtain optimal strategies $\bar{m}^*, \sigma^*$ by minimizing the simulated long-term cost $\mathcal{L}(\mathcal{S})$ (Eq. 4) in a 2-dimensional antigenic space (see Methods for details of the simulation, optimization procedures, and phase determination). By varying two key parameters, the cost $\phi$ associated to the

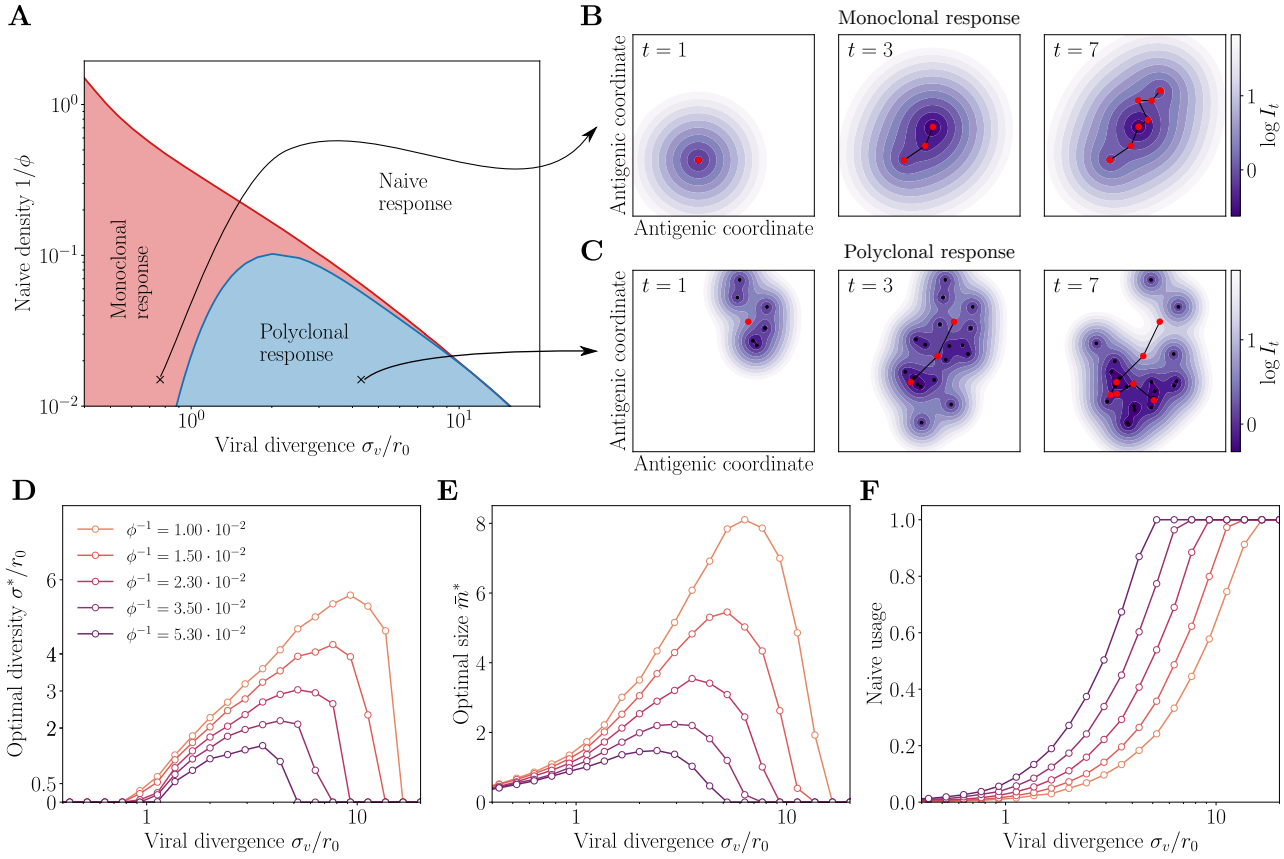


FIG. 2: **Regimes of affinity maturation.** **A.** Phase diagram of the model as a function of the naive coverage $1/\phi$, and viral divergence $\sigma_v$, in a two-dimensional antigenic map. Three phases emerge: monoclonal memory (red), polyclonal memory (purple) and naive response (white). **B-C.** Snapshots in antigenic space of the sequential immunization by a viral strain in the (B) monoclonal and (C) polyclonal phases. We show the viral position (red dots), memory clonotypes (black dots), and viral trajectory (black line). The colormap shows the log infection cost. Parameters $\sigma_v$ and $\phi$ correspond to the crosses on the phase diagram in A, with their respective optimal $\sigma^*$, $\bar{m}^*$ (see arrows). **D.** Diversity $\sigma^*$, **E.** optimal size $\bar{m}^*$, and **F.** frequency of naive cell usage in response to an immunization challenge for different naive coverages $1/\phi$. Parameters values: $\kappa = 3.3$, $\alpha = 1$, $q = 2$, $d = 2$, $\gamma = 0.85$, $\mu = 0.5$.

use of the naive repertoire, and the virus divergence $\sigma_v$, we see a phase diagram emerge with three distinct phases: the naive, monoclonal response, and polyclonal response phases (Fig. 2A). In Figs. 2B-C we show examples of the stochastic evolution of memory repertoires with optimal rules in the two phases (monoclonal and polyclonal responses). Figs. 2D-F show the behaviour of the optimal parameters, as well as the fraction of infections for which the naive repertoire is used (when the maximal infection cost $\phi$ is paid). The general shape and behaviour of this phase diagram depends only weakly on the parameter choices (see SI Fig. S1).

When the naive repertoire is sufficiently protective (small $\phi$), or when the virus mutates too much between infections (large $\sigma_v$), the optimal strategy is to produce no memory cells at all ($\bar{m}^* = 0$), and rely entirely on the naive repertoire, always paying a fixed cost $L_t = \mathcal{L} = \phi$ (naive phase).

When the virus divergence $\sigma_v$ is small relative to the cross-reactivity range $r_0$, it is beneficial to create memory clonotypes ($\bar{m}^* > 0$), but with no diversity, $\sigma^* = 0$ (monoclonal response). In this case, all newly created clonotypes are invested into a single antigenic position $a_t$ that perfectly recognizes the virus. This strategy is optimal because subsequent infections, typically caused by similar viral strains of the virus, are well recognized by these memory clonotypes.

For larger but still moderate virus divergences $\sigma_v$, this perfectly adapted memory is not sufficient to protect from mutated strains: the optimal strategy is rather to generate a polyclonal memory response, with $\bar{m}^* > 0$, $\sigma^* > 0$. In this strategy, the immune system hedges its bet against future infections by creating a large diversity of clonotypes that cover the vicinity of the encountered strain. The advantage of this policy is to anticipate future antigenic mutations of the virus; its drawback is that memories are less efficient against the current infection. This diversified pool of cells with moderate

affinity is in agreement with recent experimental observations [2, 19, 43, 44]. The diversity of the memory pool is supported by a large number of clonotypes $\bar{m}^*$ (Fig. 2D). As the virus divergence $\sigma_v$ is increased, the optimal strategy is to scatter memory cells further away from the encountered strain (increasing $\sigma^*$, Fig. 2E). However, when $\sigma_v$ it is too large, both drop to zero as the naive repertoire takes over (Fig. 2F). Increasing the naive density $\phi^{-1}$ also favors the naive phase. When there is no proliferation on average, i.e. $\Gamma \equiv \gamma(1+\mu) < 1$, there even exists a threshold $\phi_c^{-1}$ above which the naive strategy is always best (SI Fig. S1, and see Supplementary Text for estimates of that threshold).

In summary, the model predicts the two expected regimes of naive and memory use depending on the parameters that set the costs of infections and memory formation. But in addition, it shows a third phase of polyclonal response, where affinity maturation acts as an anticipation mechanism whose role is to generate a large diversity of cells able to respond to future challenges. The prediction of a less focused and thus weaker memory pool observed experimentally is thus rationalized as a result of a bet-hedging strategy.

### C. Analytical results in a solvable model

To gain insight into the transitions observed in the phase diagram of Fig. 2, we can make a series of simplifications and approximations about the model that allow for analytical progress. We assume a step function for the cross-reactivity function $f(x,a) = 1$ for $\|x-a\| \leq r_0$, and 0 otherwise, corresponding the limit $q = \infty$. Likewise, we assume a uniform distribution of viral antigenic mutations $a_{t+1} = a_t + \sigma'_v \eta'_{t+1}$, where $\eta'_t$ is a random point of the $d$-dimensional unit ball, with $\sigma'_v = \sigma_v\sqrt{1+2/d}$ (so that the variance is the same as in the Gaussian case), and similarly for memory diversification, with new clonotypes drawn from a uniform distribution is a ball of radius $\sigma' = \sigma\sqrt{1+2/d}$. The infection cost is approximated by an all-or-nothing function, with $I_t = 0$ if there is any coverage $C(a_t) > 0$, and $I_t = \phi$ if $C(a_t) = 0$. We further assume $\gamma = 0$: all clonotypes are discarded at each time step, so that memory may only be used once.

In this simplified version of the model, the phase diagram and optimal parameters can be computed analytically. One can show (see Supplementary Text) that the transition from monoclonal to polyclonal response occurs exactly when the radius of the ball within which viral mutations occur reaches the cross-reactivity radius $r_0$:

$$\sigma'_v = r_0. \tag{6}$$

Below this transition ($\sigma'_v < r_0$), the optimal strategy is to have no diversity at all and perfectly target the recognized antigen $a_t$, $\sigma'^* = 0$, as any memory cell at $a_t$ will recognize the next infection. In this case the optimal mean number of memories $\bar{m}^* = \ln(\phi/\kappa)$ results in a trade off between the cost of new memories with the risk of not developing any memory at all by minimizing $\phi e^{-\bar{m}} + \kappa\bar{m}$. The transition from the monoclonal response to naive phases is then given by $\phi = \kappa$, where $\bar{m}^* = 0$.

The polyclonal-to-naive transition may also be understood analytically. In the polyclonal response phase, the optimal strategy is, in either of the limits $\sigma'_v \gg r_0$ or $\bar{m}^* \ll 1$ (see Supplementary Text):

$$\sigma'^* \approx \sigma'_v - r_0 \tag{7}$$

$$\bar{m}^* \approx \frac{\sigma'^d_v}{r_0^d} \ln\left(\frac{\phi}{\kappa}\frac{r_0^d}{\sigma'^d_v}\right). \tag{8}$$

In particular this result becomes exact at the transition from polyclonal to naive, where $\bar{m}^* = 0$. The transition is thus given by:

$$\phi^{-1} = \frac{r_0^d}{\kappa\sigma'^d_v}. \tag{9}$$

The polyclonal response is outcompeted by the naive one when the density of naive cell ($\phi^{-1}$) becomes larger than the probability density of new strains falling within the cross-reactivity radius ($r_0^d/{\sigma'_v}^d$), rescaled by the memory cost coefficient $\kappa^{-1}$.

Fig. 3 shows the resulting phase diagram, as well as the optimal diversity $\sigma'^*$ and predicted costs for a fixed $\phi$ and $d = 1$. These predictions reproduce the main features of the full model, in particular the scaling of the immune diversity $\sigma$ with $\sigma_v$ (Fig. 2D vs. Fig. 3B) and the general shape of the optimal memory size $\bar{m}^*$ (Fig. 2E vs. Fig. 3C), which first increases as the virus becomes more divergent, to later drop to zero as memory becomes too costly to maintain and the system falls into the naive phase.

### D. Population dynamics of optimized immune systems

We now return to the full model to study the population dynamics of the memory repertoire. When the virus drifts slowly in antigenic space (small $\sigma_v$), the same clonotypes get reactivated multiple times, causing their proliferation, provided that $\Gamma = \gamma(1+\mu) > 0$. This reactivation continues until the virus leaves the cross-reactivity range of the original clonotype, at which point the memory clone decays and eventually goes extinct (Fig. 4A). Typical clonotype size trajectories from the model are shown in Fig. 4B. They show large variations in both their maximal size and lifetime. The distribution of clonotype abundances, obtained from a large number of simulations, is indeed very broad, with a power-law tail (Fig. 4C). The lifetime of clonotypes, defined as the time from emergence to extinction, is distributed according to an exponential distribution (Fig. 4D). The exponents governing the tails of these distributions, $\beta$ and $\gamma$, depend on the model parameters, in particular the divergence $\sigma_v$.

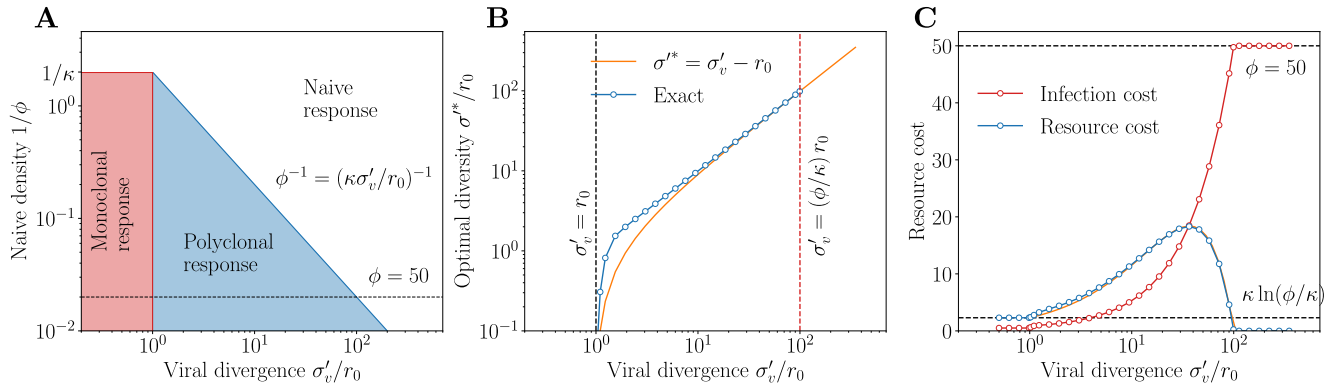


FIG. 3: **Analytical solution of a tractable model. A.** Exact phase diagram in $d = 1$ for the simplified model ($q = \infty$, $\gamma = 0$, and all-or-nothing infection cost). The boundary between monoclonal is given by $\sigma_v' = r_0$ and the boundary between polyclonal by $\phi^{-1} = (\kappa\sigma_v'/r_0)^{-d}$. **B.** Optimal memory diversity $\sigma'^* \approx \sigma_v'$ and **C.** optimal infection and plasticity costs for as a function of $\sigma_v'$ for $\phi = 50$. $\sigma'$ and $\sigma_v'$ are rescaled versions of the diversity and divergence to match the variances of the original model.

We can understand the emergence of these distributions using a simple scaling argument. After a memory clonotype has been created at $a_0$ at time 0, its size follows the approximate dynamics:

$$n_{t+1} \approx n_t\gamma\left[1 + \mu f(x, a_t)\right] = n_t\gamma\left[1 + \mu e^{-(r_t/r_0)^q}\right] \quad (10)$$

where $r_t = \|a_t - a_0\|$, the distance of the virus to the original strain, follows a random walk:

$$\langle r_t^2\rangle = t\sigma_v^2. \quad (11)$$

In (10) we have neglected demographic noise arising from the stochastic nature of duplications and deletions. As the virus drifts away and $r_t$ increases, the net growth factor $\gamma(1+\mu f)$ decreases, and eventually becomes lower than 1, for

$$r > r^* \equiv r_0 \ln\left(\frac{\gamma\mu}{1-\gamma}\right)^{1/q}. \quad (12)$$

The time it takes for this condition to be fulfilled, $t^*$, is approximately given by the diffusion law, $\langle t^*\rangle \sim (r^*/\sigma_v)^2$, assuming $\sigma_v \ll r^*$. In a very coarse approximation, the clonotype grows with rate $\Gamma = \gamma(1+\mu)$ until $t^*$, reaching peak abundance $n^* \equiv n_{t^*} \sim \Gamma^{t^*}$, and decays with rate $\gamma$ for $t > t^*$.

The tail of the distribution is dominated by rare events when the virus mutates less than expected between infections, leading to larger episodes of growth. One can show (see Supplementary Text) that the distribution of these exceptionally long $t^*$ has an exponential tail:

$$P(t^* > t) \sim e^{-t/t_s}, \quad t_s \sim \langle t^*\rangle \sim \frac{r^{*2}}{\sigma_v^2}. \quad (13)$$

This translates into a power-law tail for the peak clonotype abundance,

$$p(n^*) \sim \frac{1}{n^{*1+\beta}}, \quad \text{with } \beta \sim \frac{\sigma_v^2}{r^{*2}\ln\Gamma}. \quad (14)$$

The same scaling holds for the distribution of all abundances, since the peak determines the rest of the trajectory.

Within the same simplified picture, the lifetime $t_l$ of a clonotype is the sum of the time it takes to reach the peak, $t^*$, and the decay time until extinction, which is approximately $\ln(n^*)/\ln(1/\gamma)$:

$$t_l = t^* + \frac{\ln(n^*)}{\ln(1/\gamma)} = \left(1 + \frac{\ln\Gamma}{\ln(1/\gamma)}\right)t^*. \quad (15)$$

Thus, $t_l$ is proportional to $t^*$, and therefore also exponentially distributed:

$$p(t_l) \sim e^{-\lambda t_l}, \quad \lambda \sim \frac{\sigma_v^2}{r^{*2}}\left(1 + \frac{\ln\Gamma}{\ln(1/\gamma)}\right)^{-1}. \quad (16)$$

This simple scaling argument predicts the exponents $\beta$ and $\lambda$ fairly well: Figs. 4E-F confirm the validity of the scaling relations (14)-(16) against direct evaluation from simulations, for $d = 2$ and $q = 2$. These scalings still hold for different parameter choices (see SI Fig. S2).

These scaling relations are valid up to a geometry-dependent prefactor, which is governed by dimensionality and the shape of the cross-reactivity kernel. In the Supplementary Text, we calculate this prefactor in the special case of an all-or-nothing cross-reactivity function, $q = \infty$. Generally, $\beta$ increases with $d$, as shown in the insets of Figs. 4E-F for $q = 2$. In higher dimensions, there are more routes to escape the cross-reactivity range, and thus a faster decaying tail of large clonotypes. This effect cannot be explained by having more dimensions in which to mutate, since the antigenic variance is distributed across each dimension, according to $\sigma_v^2/d$. Rather, it results from the absence of antigenic back-mutations: in high dimensions, each mutation drifts away from the original strain with a low probability of return, making it easier for the virus to escape, and rarer for memory clonotypes to be recalled upon infections by mutant strains.

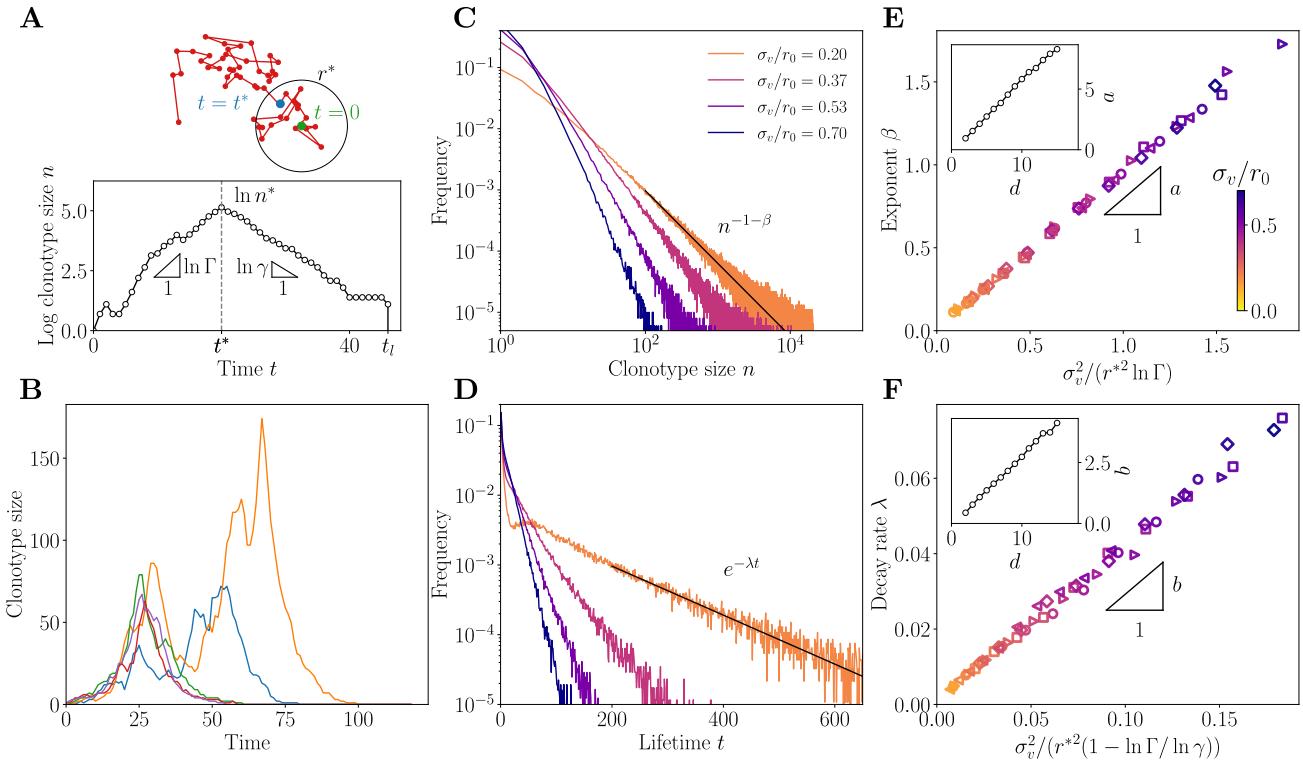


FIG. 4: **Clonotype dynamics and distribution. A.** Sketch of a recall response generated by sequential immunization with a drifting strain. Clonotypes first grow with multiplicative rate $\Gamma = \gamma(1 + \mu)$, until they reach the effective cross-reactivity radius $r^*$, culminating at $n^*$, after which they decay with rate $\gamma$ until extinction at time $t_l$. **B.** Sample trajectories of clonotypes generated by sequential immunization with a strain of mutability $\sigma_v/r_0 = 0.53$. **C.** Distribution of clonotype size for varying virus mutability $\sigma_v/r_0$. **D.** Distribution of the lifetime of a clonotype for varying virus mutability $\sigma_v/r_0$. From **B** to **D** the proliferation parameters are set to $\gamma = 0.85$, $\gamma = 0.5$ ie. $\Gamma = 1.275$. **E.** Scaling relation of the power law exponent for varying values of the parameters. Inset: dependence of the proportionality factor $a$ on dimension. **F.** Scaling relation of the decay rate $\lambda$ for varying $\sigma_v/r_0$, with scaling of the proportionality factor $b$. In both **E-F.**, the different parameters used are $(\gamma = 0.82, \mu = 0.65)$ ie. $\Gamma = 1.353$ (diamonds), $(\gamma = 0.8, \mu = 0.62)$ ie. $\Gamma = 1.296$ (squares), $(\gamma = 0.85, \mu = 0.5)$ ie. $\Gamma = 1.275$ (circles), $(\gamma = 0.87, \mu = 0.4)$ ie. $\Gamma = 1.21$ (triangles >), $(\gamma = 0.9, \mu = 0.35)$ ie. $\Gamma = 1.21$ (triangles <). From **B** to **F**, the strategy was optimized for $\phi = 100$ and $\kappa = 0.5/(1 - \gamma)$. The color code for $\sigma_v/r_0$ is consistent across the panels **C** to **F**. In this panel, the other parameters used are $\alpha = 1$, $q = 2$, $d = 2$.

### E. Comparison to experimental clone-size distributions

The power-law behaviour of the clone-size distribution predicted from the model (Fig. 4E) is in qualitative agreement with existing data. Power laws have been widely observed in immune repertoires: from early studies of repertoire sequencing data of BCR in zebrafish [5, 46], to the distribution of clonal family sizes of human IgG BCR [14, 47], as well as in T-cell receptor repertoires [6]. However, these power laws have not yet been reported in the clonotype abundance distribution of human BCR.

To fill this gap, we used publicly available IgG repertoire data of 9 human donors from a recent ultra-deep repertoire profiling study of immunoglobulin heavy-chains (IGH) [45]. The data was downloaded from Sequence Read Archive and processed as in [47]. Repertoires were obtained from the sequencing of IGH mRNA molecules to which unique molecular identifiers (UMI) were appended. For each IGH nucleotide clonotype found in the dataset, we counted the number of distinct UMI attached to it, as a proxy for the overall abundance of that clonotype in the expressed repertoire. The distributions of these abundances are shown for all 9 donors in Fig. 5A. In agreement with the theory, they display a clear power-law decay $p(n) \sim n^{-1-\beta}$, with $\beta = 1.2$-$2.4$.

Since the experimental distribution is derived from small subsamples of the blood repertoire, the absolute abundances cannot be directly compared to those of the model. In particular, subsampling means that the experimental distribution focuses on the very largest clonotypes. Thus, comparisons between model and data should be restricted to the tail behavior of the distribution, namely on its power-law exponent $\beta$.

We used this comparison to predict from the exponent $\beta$ the virus divergence between infections. To do so, we fit a linear relationship to the inset of Fig. 4E, and invert it for various values of the dimension $d$ to

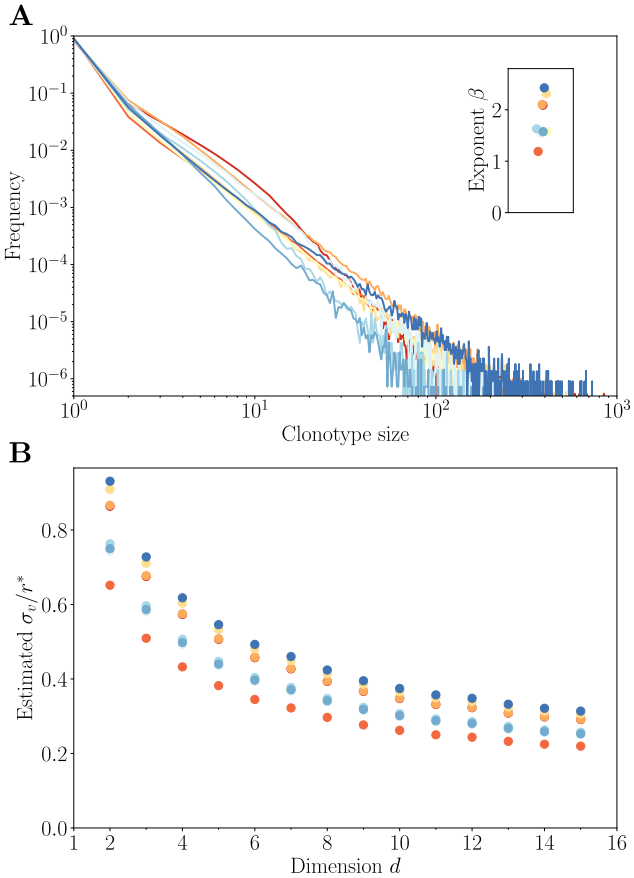


FIG. 5: **Comparison to repertoire data.** **A.** Clonotype abundance distribution of IgG repertoires of healthy donors from [45]. **B.** Estimated mutatibility $\sigma_v$ in units of the rescaled cross-reactivity $r^*$, defined as the antigenic distance at which clonotypes stop growing. $\sigma_v$ is obtained as a function of $d$ by inverting the linear relationship estimated in the inset of Fig. 4E, assuming $q = 2$ and $\Gamma = 1.4$ (estimated from [19]).

obtain $\sigma_v/r^*$. We fixed $\Gamma = 1.4$, which corresponds to a 40% boost of memory B cells upon secondary infection, inferred from a 4-fold boost following 4 sequential immunizations reported in mice [19]. The result is robust to the choice of donor, but decreases substantially with dimension, because higher dimensions mean faster escape, and thus a lower divergence for a given measured exponent (Fig. 5B). The inferred divergence $\sigma_v$ is always lower than, but of the same order as, the effective cross-reactivity range $r^*$, suggesting that the operating point of the immune system falls in the transition region between the monoclonal and polyclonal response phases (Fig. 2A).

### F. Inhibition of affinity maturation and antigenic imprinting

So far we have assumed that the immune system always activates germinal centers to implement affinity maturation, regardless of the level of protection from existing memory. However it may be optimal for the system to refrain from creating new memory cells to save costs, when the memory repertoire already responds efficiently to the current challenge. Such an approach allows the immune system to be greedy: if the early memory response induced by existing cells is fast enough, it can inhibit the maturation of the repertoire.

We introduce this inhibition of affinity maturation in our model through a cost threshold $0 \leq \xi \leq \phi$ below which the formation of new clonotypes is suppressed. When $L_t < \xi$, the memory repertoire recognizes the viral challenge with sufficient efficacy to offer protection, and the immune system decides to rely on existing clonotypes rather than to create new ones, saving the plasticity cost $\kappa m$. However, existing clonotypes are expanded, or "backboosted," proportionally to their affinity to the new infection as before. The threshold $\xi$ is left as an optimization parameter, in addition to $\sigma$ and $\bar{m}$. In the special case $\xi = 0$, we recover the original model where affinity maturation always occurs.

We find that there exists an optimal threshold $\xi^* > 0$ in both the polyclonal and monoclonal response phases. The frequency of affinity maturation events mostly depends on $\sigma_v$ (Fig. 6A). While this threshold remains approximately constant, the frequency of affinity maturation events increases as $\sigma_v$ increases. At small $\sigma_v$, the optimal strategy is to extensively backboost existing memory cells; for large $\sigma_v$, the growing unpredictability of the next viral move makes it more likely to have recourse to affinity maturation. In other words, when the virus is stable (low $\sigma_v$), the immune system is more likely to capitalize on existing clonotypes, and not implement affinity maturation, because savings made on the plasticity cost outweigh the higher infection cost. As the virus drifts away with time, this infection cost also increases, until it reaches the point where affinity maturation becomes worthwile again.

Typical trajectories of the infection cost show the typical dynamics induced by backboosting, with long episodes where existing memory remains sufficient to keep the cost below $\xi$ (Fig. 6B), interrupted by infections that fall too far away from existing memory, triggering a new episode of affinity maturation and concomitant drop in the infection cost. This effect is known as "antigenic imprinting," and is linked to the notion of "original antigenic sin," whereby the history of past infections determines the process of memory formation, usually by suppressing it. This imprinting leads to the couterintuitive prediction that a better experienced immune system is less likely to form efficient memory upon new infections.

We call the time between affinity maturation events $t_\xi$. Its mean $\langle t_\xi \rangle$ is equal to the inverse of the frequency

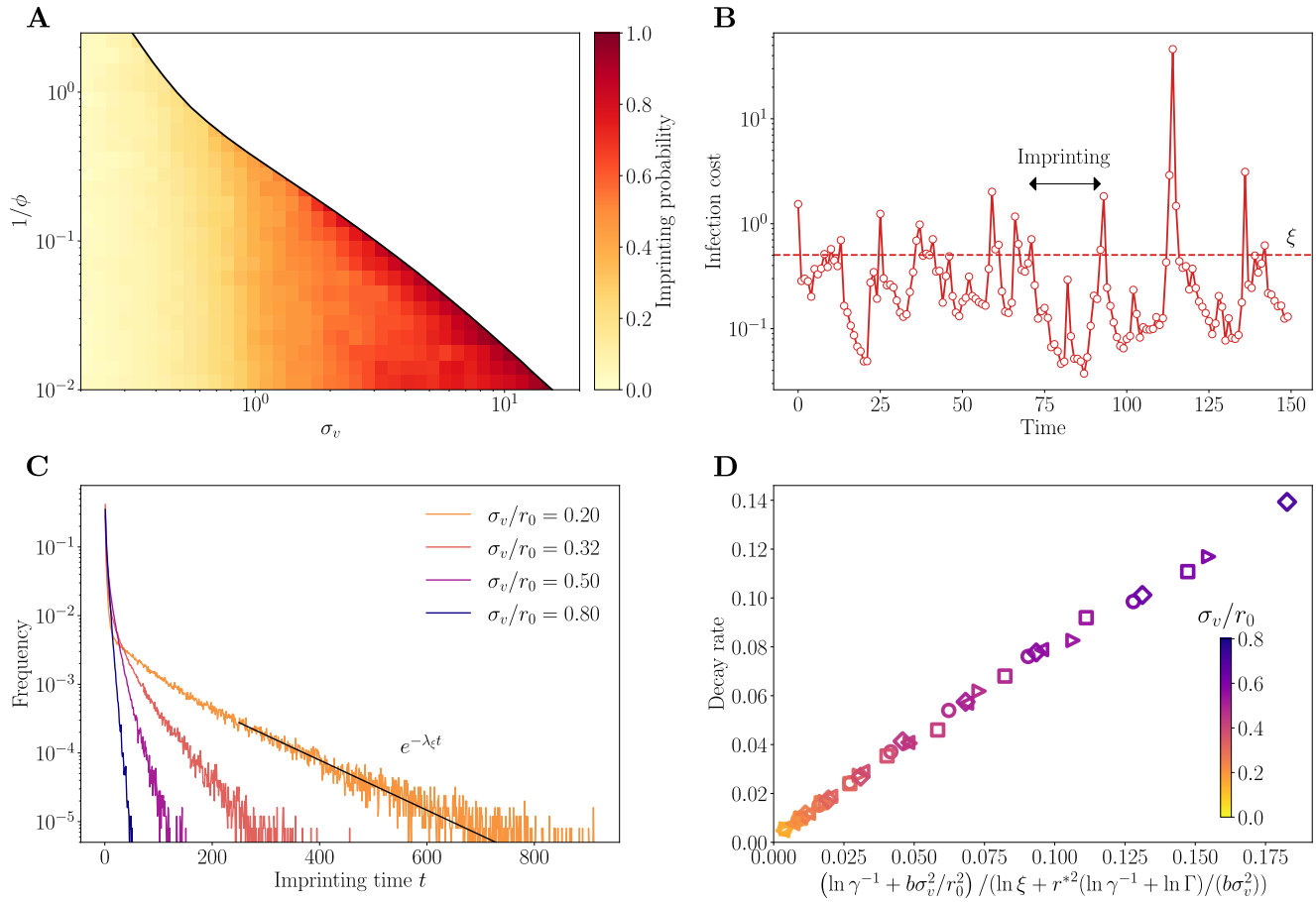


FIG. 6: **Imprinting and backboosting.** Affinity maturation is now made optional depending on the cost of infections. **A.** Frequency of infections leading to affinity maturation in the optimal strategy. The frequency increases with the virus divergence $\sigma_v$, up to the point where the transitions to the naive phase where memory is not used at all. **B.** Typical trajectory of infection cost in sequential infections at $\sigma_v/r_0 = 0.5$. When the cost goes beyond the threshold $\xi$, affinity maturation is activated, leading to a drop in infection cost. These periods of sub-optimal memory describe an "original antigenic sin," whereby the immune system is frozen in the state imprinted by the last maturation event. **C.** The distribution of imprinting times, i.e. the number of infections between affinity maturation events, decays exponentially. The proliferation parameters in **A** to **C** are set to $\gamma = 0.85$ and $\mu = 0.5$. **D.** Predicted scaling with the model parameters of the decay rate $\lambda_\xi$ of the distribution of impriting times $P(t_\xi) \sim e^{-\lambda_\xi t_\xi}$ from **C**. In **D**, the different parameters used are ($\gamma = 0.82, \mu = 0.65$) ie. $\Gamma = 1.353$ (diamonds), ($\gamma = 0.8, \mu = 0.62$) ie. $\Gamma = 1.296$ (squares), ($\gamma = 0.85, \mu = 0.5$) ie. $\Gamma = 1.275$ (circles), ($\gamma = 0.87, \mu = 0.4$) ie. $\Gamma = 1.21$ (triangles >), ($\gamma = 0.9, \mu = 0.35$) ie. $\Gamma = 1.21$ (triangles <). The color code for $\sigma_v/r_0$ is consistent across the panels **C** and **D**. From **B** to **D**, the strategy was optimized for $\phi = 100$ and $\kappa = 0.5/(1-\gamma)$. In this panel, the other parameters used are $\alpha = 1$, $q = 2$, $d = 2$.

of maturation events, and thus decreases with $\sigma_v$. Its distribution, shown Fig. 6C, has an exponential tail with exponent $\lambda_\xi$. Affinity maturation should happen when the distance of the virus to the original strain, $r_t$, has become too large,

$$n_t e^{-(r_t/r_0)^q} < \xi^{-1/\alpha}. \tag{17}$$

The exponential tail of the distribution of $t_\xi$ is dominated by episodes where the viral strain drifted less than expected. In that case, the originally matured clonotype grows to a large size, offering protection for a long time, even after it has stopped growing, when $t > t^*$. Using the approximate behaviour of the clonotype size during this decay $n_t \sim n^* \gamma^{t-t^*}$ (Fig. 4A), this gives the condition:

$$\frac{r_t^q}{r_0^q} - t^* \ln\Gamma + (t - t^*) \ln\frac{1}{\gamma} > \frac{\ln\xi}{\alpha} \tag{18}$$

Past the critical radius $r^*$, clonotype sizes start shrinking ($t\ln(1/\gamma)$ term), and the virus also drifts further away ($r_t^q$ term). Both effects contribute to crossing the threshold. Since we focus on the tail of the distribution, we may replace $t^*$ by $r^*/b\sigma_v^2$, and $r_t^2$ by $bt\sigma_v^2$, where $b$ is the scaling factor determining the decay exponent of the first passage time at a given radius. Doing so yields the following

prediction for the decay rate:

$$\lambda_\xi \sim \frac{\ln(1/\gamma) + b\sigma_v^2/r_0^2}{\ln\xi + r^{*2}(\ln(1/\gamma) + \ln\Gamma)/(b\sigma_v^2)}. \quad (19)$$

Fig. 6D verifies the validity of this scaling, confirming the prediction of rare long episodes of re-infections without affinity maturation.

## III. DISCUSSION

Adaptive immunity coordinates multiple components and cell types across entire organisms over many space and time scales. Because it is difficult to empirically characterize the immune system at the system scale, normative theories have been useful to generate hypotheses and fill the gap between observations at the molecular, cellular, and organismal scales [48, 49]. Such approaches include clonal selection theory [50], or early arguments about the optimal size and organization of immune repertoires [39, 51, 52], and of affinity maturation [21, 53]. While these theories do not rely on describing particular mechanisms of immune function, they may still offer quantitative insights and help predict new mechanisms or global rules of operation.

Previous work developed models of repertoire organization as a constrained optimization problem where the expected future harm of infection, or an *ad hoc* utility function, is minimized [30, 41, 42, 54]. In Ref. [42], it was assumed clonotypes specific to all antigens are present at all times in the repertoire; the mechanism of immune memory then merely consists of expanding specific clonotypes at the expense of others. This assumption describes T-cell repertoires well, where there are naive cells with good affinity to essentially any antigen [55]. For B cells the situation is more complex because of affinity maturation. In addition, re-organizing the repertoire through mutation and selection has a cost, and is subject to metabolic and physical constraints.

Our work addresses these challenges by proposing a framework of discrete-time decision process to describe the optimal remodeling of the B-cell repertoire following immunization, through a combination of affinity maturation and backboosting. While similar to [30], our approach retains the minimal amount of mechanistic details and focuses on questions of repertoire remodeling, dynamics, and structure.

We investigated strategies that maximize long-term protection against subsequent challenges and minimize short-term resource costs due to the affinity maturation processes. Using this model, we observed that optimal strategies may be organized into three main phases as the pathogen divergence and naive coverage are varied. For an expected divergence lower than the cross-reactivity range, we found that the optimal response is monoclonal, with a specificity perfectly matching the one of the infecting strain. As the strain divergence increases, we observed that the optimal formation of memory becomes polyclonal, with diversity of the memory scaling with the viral divergence, leading to a weakly reactive memory pool to the infecting viral strain. This optimal memory response remains better than the naive response, as long as the next challenge may be anticipated within the cost constraints of affinity maturation.

We expect these distinct phases to co-exist in the same immune system, as there exists a wide range of pathogen divergences, depending on their evolutionary speed and typical frequency of recurrence. For fast recurring or slowly evolving pathogens, the monoclonal response ensures a very specific and targeted memory. This role could be played by long-lived plasma cells. These cells are selected through the last rounds of affinity maturation, meaning that they are very specific to the infecting strain [56]. Yet, despite not being called memory cells, they survive for long times in the bone marrow, providing long-term immunity. For slow recurring or fast evolving pathogens, the polyclonal response provides a diverse memory to preempt possible antigenic drift of the pathogen. The function could be fulfilled by memory B cells, which are released earlier in the affinity maturation process, implying that they are less specific to the infecting strain, but collectively can cover more immune escape mutations. As a sidenote, there is an ongoing debate as to whether affinity maturation starts from memory or naive B cells during sequential challenges [57, 58]. Our model does not commit to either scenario since we assume that the main benefit of memory is on the infection cost, rather than its re-use in subsequent rounds of affinity maturation.

We extended our model to account for the inhibition of maturation processes in the presence of a strongly reactive memory pool. We unveiled a trade-off between the exploitation of existing reactive memory cells through the inhibition of the affinity maturation processes, and the anticipation of future challenges through the generation of new memory clonotypes. We observed that the optimal immune repertoire evolution is frequently stalled for small expected viral divergence, with a strong inhibition of affinity maturation for subsequent challenges. Such an observation can be interpreted as the phenomenon of antigenic imprinting widely described in sequential immunization assays [35], or "original antigenic sin" [34]. This effect implies that having been exposed to previous strains of the virus is detrimental to mounting the best possible immune response. This counter-intuitive strategy is rationalized within our optimization framework: lack of affinity maturation happens only when previous memories are protective enough (relative to the extra cost $\kappa\bar{m}$ of creating new specificities), and does not happen at all in naive individuals.

While we considered a single antigenic site, our results generalize to multiple immune targets. We expect such generalization to create competition and trade-offs between different antigens, and perhaps different regimes describing whether the immune system tries to target multiple antigens or instead focuses on a single one (im-

munodominance). For instance, while our interpretation of the original antigenic sin involves a single epitope, a similar concept exists for multiple epitopes [32]. When during sequential challenges only one of multiple epitopes changes at a time, it may be optimal for the immune system to rely on its protections against the invariant epitopes. Only after all epitopes have escaped immunity does affinity maturation get re-activated concomitantly to a spike of infection harm, similar to our result for a single antigen.

We only considered a single pathogenic species at a time, with the assumption that pathogens are antigenically independent, so that the costs relative to different pathogens are additive. Possible generalizations could include secondary infections, as well as antigenically related pathogens showing patterns of cross-immunity (such as cowpox and smallpox, or different groups of influenza), which could help us shed light on complex immune interactions between diseases and serotypes, such as negative interference between different serotypes of the Dengue fever leading to hemorrhagic fever, or of the human Bocavirus affecting different body sites or [34].

## IV. MATERIALS AND METHODS

The average cumulative cost $\mathcal{L}$ in Eq. 4 is approximated by a Monte-Carlo method. To ensure the simulated repertoire reaches stationarity, we start from a naive repertoire and discard an arbitrary number of initial viral encounters. Because the process is ergodic, simulating a viral-immune trajectory over a long time horizon is equivalent to simulating $M$ independent trajectories of smaller length $T$. To ensure the independence of our random realizations across our $M$ parallel streams we use the random number generators method *split* provided in Tina's RNG library [59]. The cumulative cost function $\mathcal{L}$ is convex for the range of parameters tested. To optimize $\mathcal{L}$ under positivity constraints for the objective variables $\sigma$, $\bar{m}$ and $\xi$, we use Py-BOBYQA [60], a coordinate descent algorithm supporting noisy objective functions.

The polyclonal to monoclonal (red curve) and memory to naive (blue curve) boundaries of the phase diagrams in Fig. 3A and Fig. S2 are obtained by respectively solving $\partial\mathcal{L}/\partial\sigma = 0$ in the monoclonal phase and $\partial\mathcal{L}/\partial\bar{m} = 0$ in the naive phase. Both these derivatives can be approximated by finite differences with arbitrary tolerances on $\sigma$ and $\bar{m}$. We fix the tolerance on $\sigma$ to 0.2 and the tolerance on $\bar{m}$ to 0.01. To obtain the root of these difference functions, we use a bisection algorithm. In order to further decrease the noise level, we compute the difference functions across pairs of simulations, each pair using an independent sequence of pathogens $a_t$ of length $L = 400$. The number of independent pairs of simulations used for each value of $\sigma_v$ and $\phi$ is $M \sim 10^5$.

**Acknowledgements**

The study was supported by the European Research Council COG 724208 and ANR-19-CE45-0018 "RESP-REP" from the Agence Nationale de la Recherche and DFG grant CRC 1310 "Predictability in Evolution". The authors thank Natanael Spisak for pre-processing the raw data from [45], and for useful discussions and suggestions.

[1] Nieuwenhuis P, Opstelten D (1984) Functional anatomy of germinal centers. *American Journal of Anatomy* 170:421–435.

[2] Victora GD, Nussenzweig MC (2012) Germinal Centers. *Annual Review of Immunology* 30:429–457.

[3] Eisen HN, Siskind GW (1964) Variations in Affinities of Antibodies during the Immune Response *. *Biochemistry* 3:996–1008.

[4] Weisel F, Shlomchik M (2017) Memory B Cells of Mice and Humans. *Annual Review of Immunology* 35:255–284.

[5] Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR (2009) High-Throughput Sequencing of the Zebrafish Antibody Repertoire. *Science* 324:807–810.

[6] Mora T, Walczak AM (2019) in *Systems Immunology* (CRC Press).

[7] Kocks C, Rajewsky K (1988) Stepwise intraclonal maturation of antibody affinity through somatic hypermutation. *Proceedings of the National Academy of Sciences* 85:8206–8210.

[8] Kleinstein SH, Louzoun Y, Shlomchik MJ (2003) Estimating Hypermutation Rates from Clonal Tree Data. *The Journal of Immunology* 171:4639–4649.

[9] Kepler TB (2013) Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000Research* 2:103.

[10] Yaari G, Kleinstein SH (2015) Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Medicine* 7:121.

[11] Ralph DK, Iv FAM (2016) Likelihood-Based Inference of B Cell Clonal Families. *PLOS Computational Biology* 12:e1005086.

[12] Liao HX, et al. (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496:469–476.

[13] Nourmohammad A, Otwinowski J, Łuksza M, Mora T, Walczak AM (2019) Fierce Selection and Interference in B-Cell Repertoire Response to Chronic HIV-1. *Molecular Biology and Evolution* 36:2184–2194.

[14] Jiang N, et al. (2013) Lineage Structure of the Human Antibody Repertoire in Response to Influenza Vaccination. *Science Translational Medicine* 5:171ra19–171ra19.

[15] Horns F, Vollmers C, Dekker CL, Quake SR (2019) Signatures of selection in the human antibody repertoire: Selective sweeps, competing subclones, and neutral drift. *Proceedings of the National Academy of Sciences* 116:1261–1266.

[16] Lee J, et al. (2016) Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nature Medicine* 22:1456–1464.

[17] Smith KGC, Light A, Nossal GJV, Tarlinton DM (1997) The extent of affinity maturation differs between the memory and antibody-forming cell compartments in the primary immune response. *The EMBO Journal* 16:2996–3006.

[18] Weisel FJ, Zuccarino-Catania GV, Chikina M, Shlomchik MJ (2016) A Temporal Switch in the Germinal Center Determines Differential Output of Memory B and Plasma Cells. *Immunity* 44:116–130.

[19] Viant C, et al. (2020) Antibody Affinity Shapes the Choice between Memory and Germinal Center B Cell Fates. *Cell* 183:1298–1311.e11.

[20] Oprea M, Perelson AS (1997) Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts. *The Journal of Immunology* 158:5155–5162.

[21] Oprea M, Van Nimwegen E, Perelson AS (2000) Dynamics of one-pass germinal center models: Implications for affinity maturation. *Bulletin of Mathematical Biology* 62:121–153.

[22] Wang S, et al. (2015) Manipulating the Selection Forces during Affinity Maturation to Generate Cross-Reactive HIV Antibodies. *Cell* 160:785–797.

[23] Sachdeva V, Husain K, Sheng J, Wang S, Murugan A (2020) Tuning environmental timescales to evolve and maintain generalists. *Proceedings of the National Academy of Sciences* 117:12693–12699.

[24] Molari M, Eyer K, Baudry J, Cocco S, Monasson R (2020) Quantitative modeling of the effect of antigen dosage on B-cell affinity distributions in maturating germinal centers. *eLife* 9:e55678.

[25] Grenfell BT (2004) Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science* 303:327–332.

[26] Blanquart F, Gandon S (2013) Time-shift experiments and patterns of adaptation across time and space. *Ecology Letters* 16:31–38.

[27] Cobey S, Wilson P, Matsen FA (2015) The evolution within us. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370:20140235.

[28] Koelle K, Cobey S, Grenfell B, Pascual M (2006) Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans. *Science* 314:1898–1903.

[29] Marchi J, Lässig M, Walczak AM, Mora T (2021) Antigenic waves of virus–immune coevolution. *Proceedings of the National Academy of Sciences* 118:e2103398118.

[30] Schnaack OH, Nourmohammad A (2021) Optimal evolutionary decision-making to store immune memory. *eLife* 10:e61346.

[31] Francis T (1960) On the Doctrine of Original Antigenic Sin. *Proceedings of the American Philosophical Society* 104:572–578.

[32] Cobey S, Hensley SE (2017) Immune history and influenza virus susceptibility. *Current Opinion in Virology* 22:105–111.

[33] Hoskins TW, Davies J, Smith AJ, Miller C, Allchin A (1979) Assessment of inactivated influenza-A vaccine after three outbreaks of influenza A at Christ's Hospital. *The Lancet* 313:33–35.

[34] Vatti A, et al. (2017) Original antigenic sin: A comprehensive review. *Journal of Autoimmunity* 83:12–21.

[35] Kim JH, Skountzou I, Compans R, Jacob J (2009) Original Antigenic Sin Responses to Influenza Viruses. *The Journal of Immunology* 183:3294–3301.

[36] Linderman SL, Hensley SE (2016) Antibodies with 'Original Antigenic Sin' Properties Are Valuable Components of Secondary Immune Responses to Influenza Viruses. *PLOS Pathogens* 12:e1005806.

[37] Yewdell JW, Santos JJS (2020) Original Antigenic Sin: How Original? How Sinful? *Cold Spring Harbor Perspectives in Medicine* p a038786.

[38] Worobey M, Plotkin S, Hensley SE (2020) Influenza Vaccines Delivered in Early Childhood Could Turn Antigenic Sin into Antigenic Blessings. *Cold Spring Harbor Perspectives in Medicine* p a038471.

[39] Perelson AS, Oster GF (1979) Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. *Journal of Theoretical Biology* 81:645–670.

[40] Cohen C, et al. (2021) Asymptomatic transmission and high community burden of seasonal influenza in an urban and a rural community in South Africa, 2017–18 (PHIRST): A population cohort study. *The Lancet Global Health* 9:e863–e874.

[41] Mayer A, Balasubramanian V, Mora T, Walczak AM (2015) How a well-adapted immune system is organized. *Proceedings of the National Academy of Sciences* 112:5950–5955.

[42] Mayer A, Balasubramanian V, Walczak AM, Mora T (2019) How a well-adapting immune system remembers. *Proceedings of the National Academy of Sciences* 116:8815–8823.

[43] Tas JMJ, et al. (2016) Visualizing antibody affinity maturation in germinal centers. *Science* 351:1048–1054.

[44] Kuraoka M, et al. (2016) Complex Antigens Drive Permissive Clonal Selection in Germinal Centers. *Immunity* 44:542–552.

[45] Briney B, Inderbitzin A, Joyce C, Burton DR (2019) Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* 566:393–397.

[46] Mora T, Walczak AM, Bialek W, Callan CG (2010) Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences* 107:5405–5410.

[47] Spisak N, Walczak AM, Mora T (2020) Learning the heterogeneous hypermutation landscape of immunoglobulins from high-throughput repertoire data. *Nucleic Acids Research* 48:10702–10712.

[48] Chakraborty AK (2017) A Perspective on the Role of Computational Models in Immunology. *Annual Review of Immunology* 35:403–439.

[49] Altan-Bonnet G, Mora T, Walczak AM (2020) Quantitative Immunology for Physicists. *Physics Reports* 849:1–83.

[50] Burnet FM (1957) A modification of Jerne's theory of antibody production using the concept of clonal selection. *The Australian Journal of Science* 20:67–69.

[51] Perelson AS, Mirmirani M, Oster GF (1976) Optimal strategies in immunology. I. B-cell differentiation and proliferation. *Journal of Mathematical Biology* 3:325–367.

[52] Perelson AS, Mirmirani M, Oster GF (1978) Optimal strategies in immunology. II. B memory cell production. *Journal of Mathematical Biology* 5:213–256.

[53] Kepler TB, Perelson AS (1993) Somatic Hypermutation

in B Cells: An Optimal Control Treatment. *Journal of Theoretical Biology* 164:37–64.
[54] Marsland R, Howell O, Mayer A, Mehta P (2021) Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response. *Proceedings of the National Academy of Sciences* 118:e2011709118.
[55] Moon JJ, et al. (2007) Naive CD4+ T Cell Frequency Varies for Different Epitopes and Predicts Repertoire Diversity and Response Magnitude. *Immunity* 27:203–213.
[56] Akkaya M, Kwak K, Pierce SK (2020) B cell memory: Building two walls of protection against pathogens. *Nature Reviews Immunology* 20:229–238.
[57] Pape KA, Jenkins MK (2018) Do Memory B Cells Form Secondary Germinal Centers? It Depends. *Cold Spring Harbor Perspectives in Biology* 10:a029116.
[58] Shlomchik MJ (2018) Do Memory B Cells Form Secondary Germinal Centers? Yes and No. *Cold Spring Harbor Perspectives in Biology* 10:a029405.
[59] Bauke H, Mertens S (2007) Random numbers for large-scale distributed Monte Carlo simulations. *Physical Review E* 75:066701.
[60] Cartis C, Fiala J, Marteau B, Roberts L (2019) Improving the Flexibility and Robustness of Model-based Derivative-free Optimization Solvers. *ACM Transactions on Mathematical Software* 45:32:1–32:41.
[61] Redner S (2001) *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge).

## Supplementary information

### Appendix A: Mean-field naive coverage

Here we show how the infection cost function defined in the main text,

$$I_t = \min\left[\phi, \left(\sum_{x\in P_{t-1}} c_{x,t} f(x,a_t)\right)^{-\alpha}\right], \tag{A1}$$

may be derived as the mean-field limit of a repertoire with memory and naive compartments.

In addition to the evolving memory repertoire $P_t$ already described in the main text, we define a naive repertoire made of random receptors $\mathcal{N}$, distributed uniformly with density $\rho$. Viruses may be recognized by either the memory or naive clonotypes. The naive coverage is defined as:

$$C_{\text{naive}}(a_t) = \sum_{x\in\mathcal{N}} f(x,a_t), \tag{A2}$$

and the memory coverage as before:

$$C(a_t) = \sum_{x\in P_{n-1}} c_{x,t} f(x,a_t). \tag{A3}$$

(In this convention, each naive clonotype has size one in arbitrary units.)

Depending on the values of these coverages, the system will choose to use either the naive repertoire, or an existing memory. In this decision, we factor in the fact that using the naive repertoire is more costly, which we account for using a prefactor $\beta < 1$. The cost is then defined as:

$$L_t = \max\left[\beta C_{\text{naive}}(a_t), C(a_t)\right]^{-\alpha}. \tag{A4}$$

We can simplify this expression in the limit where naive clonotypes are very numerous, but each offer weak coverage. In the limit of high density of naive cells, $\rho \to \infty$, the coverage self-averages to its mean value:

$$C_{\text{naive}} \approx \langle C_{\text{naive}}\rangle = \rho \int d^d x\, f(x,a_t) = \rho U_d(q) r_0^d, \tag{A5}$$

with

$$U_d(q) = \int d^d y\, e^{-\|y\|^q} = S_d \int r^{d-1} dr\, e^{-r^q}, \tag{A6}$$

where we have done the change of variable $x = a_t + y r_0$, and where $S_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the unit sphere.

Taking the $\rho \to \infty$ and $\beta \to 0$ limits, while keeping $\beta\rho$ finite, corresponds to a dense naive repertoire but where each naive cell weakly covers the antigenic space. In this limit we recover the model of the main text

$$I_t = \min\left[\phi, C(a_t)^{-\alpha}\right], \tag{A7}$$

with $\phi = (\rho\beta U_d(q) r_0^d)^{-\alpha}$.

### Appendix B: Transition from monoclonal to naive phase at $\sigma_v = 0$

Here we derive an expression for the phase boundary between the monoclonal and polyclonal phases in the limit $\sigma_v = 0$, where the virus does not move.

In the special case where $\mu = 0$, clonotypes cannot multiply. At each time step, a number $m_n$ of new clonotypes at created at $a_n = \text{const}$, distributed according to a Poisson law of mean $\bar{m}$. This number is added to existing clonotypes, of which a random fraction $\gamma$ survives. If the previous number of clonotypes, $M_n$, is Poisson distributed with mean $\bar{M}_n$, the number of surviving ones $M'_n$ is also Poisson distributed with mean $\gamma\bar{M}_n$ (since subsampling a Poisson-distributed number still gives a Poisson law). Then, the new number of clonotypes, $M_{n+1} = M'_n + m_n$, is also Poisson distributed, with the recurrence relation:

$$\bar{M}_{n+1} = \gamma\bar{M}_n + \bar{m}. \tag{B1}$$

At steady state, we have

$$\bar{M}_n \to \frac{\bar{m}}{1-\gamma}. \tag{B2}$$

Since all clonotypes are at $x = a_n$, the coverage is $C(a_n) = M$, so that the expected cost reads:

$$\mathcal{L} = \phi \exp\left(-\frac{\bar{m}}{1-\gamma}\right) + \sum_{M=1}^{+\infty} \frac{1}{M^\alpha} \exp\left(-\frac{\bar{m}}{1-\gamma}\right) \frac{1}{m!}\left(\frac{\bar{m}}{1-\gamma}\right)^M + \kappa\bar{m}. \tag{B3}$$

To find the transition from monoclonal to naive, $\bar{m} = 0$, we need to find the value of $\phi$ for which $\partial\mathcal{L}/\partial\bar{m}$ changes sign at $\bar{m} = 0$: if this derivative is positive, it is better to have $\bar{m} = 0$ (since the function is convex); if it is negative, there is benefit to be gained by increasing $\bar{m} > 0$. The condition:

$$\left.\frac{\partial\mathcal{L}}{\partial\bar{m}}\right|_{\bar{m}=0} = \kappa - \frac{\phi}{1-\gamma} + \frac{1}{1-\gamma} \tag{B4}$$

gives the transition point

$$\phi_c = 1 + \kappa(1-\gamma) \tag{B5}$$

For $\mu > 0$, we redefine $M_n$ as the sum of all clonotype sizes, which is equal to the coverage, $C = M_n = \sum_{x\in P_n} c_{x,n}$. The recurrence relation is replaced by:

$$\bar{M}_{n+1} = \gamma(1+\mu)\bar{M}_n + \bar{m}. \tag{B6}$$

For $\gamma(1+\mu) > 1$, this number explodes, so that $M$ is infinite, reducing the infection cost to 0 regardless of $\bar{m}$. The transition point is then

$$\phi_c = 0. \tag{B7}$$

For $\gamma(1+\mu) < 1$, $\bar{M}_n$ reaches a steady state value,

$$\bar{M}_n \to \frac{\bar{m}}{1-\gamma(1+\mu)}. \tag{B8}$$

Although $M_n$ is not strictly distributed according to a Poisson law, it is still a good approximation, so that we can repeat the same argument as with $\mu = 0$,

$$\phi_c \approx 1 + \kappa(1-\gamma(1+\mu)). \tag{B9}$$

### Appendix C: Solutions to the simplified model of section "Analytical results in a solvable model"

Here we derive a general expression for the simplified model defined in the second *Results* subsection of the main text in integral form. Then we derive analytically the boundaries of the phase diagram, as well as an approximate solution in the polyclonal phase. Finally we give exact analytical expressions for the 1-dimensional case.

The simplified model is defined by the following cost:

$$L_t = I_t + \kappa m_t, \tag{C1}$$

where $I_t = \phi$ if $C(a_n) = 0$, and $I_t = 0$ if $C(a_n)$. In addition, all memories die after their first use, which is realized by $\gamma = 0$.

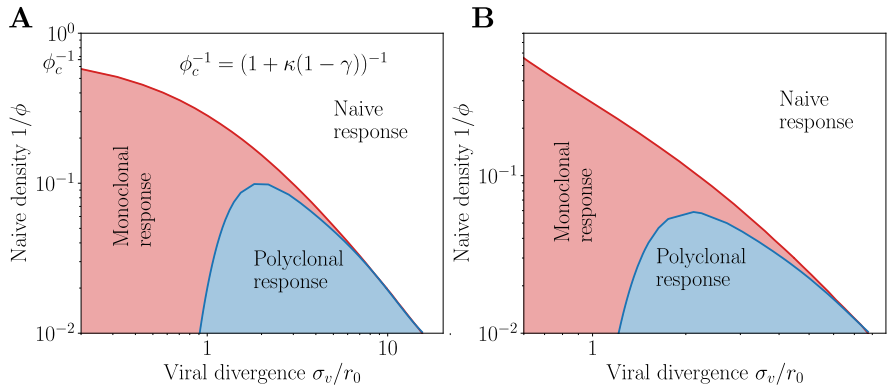


FIG. S1: **Phase diagram for various parameters. A.** Phase diagram for parameters $\mu = 0$, $\gamma = 0.85$ and $d = 2$. We observe the upper transition point $\phi_c = (1 + \kappa(1-\gamma))^{-1}$ at $\sigma_v = 0$. **B.** Phase diagram for parameters $\mu = 0.5$, $\gamma = 0.85$ and $d = 3$. Since $\gamma(1+\mu) > 1$ the transition point $\phi_c = \infty$. In both **A** and **C** we observe that the phase diagram retains the same shape. In this panel $\alpha = 1$, $q = 2$ and $\kappa = 0.5/(1-\gamma)$.

### 1. General formulation

Define $P_{\text{hit}}(\sigma', r_0, r)$ as the probability that a random memory will recognize the next infection at distance $r$, i.e. the probability that a random point in the ball of radius $\sigma'$ and a point at distance $r$ from its center are at distance $\leq r_0$. The probability that none of $m$ clonotypes recognize the virus, where $m$ is drawn from a Poisson distribution of mean $\bar{m}$, reads:

$$P_{\text{miss}}(\sigma', r_0, r) = \sum_m e^{-\bar{m}} \frac{\bar{m}^m}{m!} (1 - P_{\text{hit}}(\sigma', r_0, r))^m = e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)}. \tag{C2}$$

The best strategy maximizes this probability, averaged over the location of the next infection, over $\sigma'$:

$$\bar{P}_{\text{miss}}(\sigma', \bar{m}, r_0, \sigma'_v) = \left\langle e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)} \right\rangle_{\mathcal{B}(\sigma'_v)} = \frac{1}{\sigma_v'^d V_d} \int_0^{\sigma'_v} S_d r^{d-1} dr \, e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)} \tag{C3}$$

where $\mathcal{B}_{\sigma'_v}$ is the ball of radius $\sigma'_v$, $V_d = \pi^{d/2}/\Gamma(d/2+1)$ is the volume of a unit ball, and $S_d = 2\pi^{d/2}/\Gamma(d/2)$ the area of the unit sphere.

Then the expected overall cost reads:

$$\mathcal{L} = \phi \bar{P}_{\text{miss}}(\sigma', \bar{m}, r_0, \sigma_v) = \phi \left\langle e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)} \right\rangle_{\mathcal{B}(\sigma'_v)} + \kappa \bar{m}. \tag{C4}$$

### 2. Exact location of the phase transitions, and approximate solution in the polyclonal phase

The location of the optimal $\sigma'$ may be rigorously bounded from above and below. If $\sigma' < \sigma'_v - r_0$, then only part of the future positions of the virus are covered, so increasing $\sigma'$ can bring no harm. Likewise, for $\sigma' > \sigma'_v + r_0$, memory covers parts of the antigenic space that have no chance of harboring the next virus, so that decreasing $\sigma'$ is also always advantageous. Thus, the optimum $\sigma'^*$ must satisfy:

$$\sigma'_v - r_0 \leq \sigma'^* \leq \sigma'_v + r_0. \tag{C5}$$

As already argued in the main text, when $\sigma'_v < r_0$, there is clearly no benefit to having $\sigma' > 0$, so the optimum is reached at $\sigma' = 0$. (C5) further shows that if $\sigma'_v > r_0$, then $\sigma'^* > 0$, so that a polyclonal phase is optimal. As a consequence, the transition from the monoclonal to polyclonal phases happens exactly at

$$\text{Monoclonal to polyclonal:} \quad \sigma'_v = r_0. \tag{C6}$$

In the monoclonal phase, memory always recognizes the next virus. The only risk of paying $\phi$ is when no memory is created, which happens with probability $e^{-\bar{m}}$, so that the cost reads:

$$\mathcal{L} = \phi e^{-\bar{m}} + \kappa \bar{m}. \tag{C7}$$

The optimal $\bar{m}^* = \ln(\phi/\kappa)$ cancels at the monoclonal-to-naive transition:

$$\text{Monoclonal to naive:} \quad \phi = \kappa. \tag{C8}$$

In the polyclonal phase, we could not find a general analytical solution, but there are two limits in which the solution may be calculated. The first limit is when $\sigma_v \gg r_0$. In that case, (C5) implies $\sigma'^* \approx \sigma'_v$, and

$$P_{\text{hit}}(\sigma'^*, r_0, r) \approx \frac{r_0^d}{\sigma_v'^d}, \tag{C9}$$

which doesn't depend on $r$. Then, minimizing

$$\mathcal{L} \approx \phi e^{\bar{m} P_{\text{hit}}} + \kappa \bar{m} \tag{C10}$$

with respect to $\bar{m}$ yields:

$$\bar{m}^* = \frac{1}{P_{\text{hit}}} \ln\left(\frac{\phi}{\kappa} P_{\text{hit}}\right) = \frac{\sigma_v'^d}{r_0^d} \ln\left(\frac{\phi}{\kappa} \frac{r_0^d}{\sigma_v'^d}\right). \tag{C11}$$

The second limit in which things simplify is for small $\bar{m}$. Then the exponential in (C3) may be expanded at first order, yielding:

$$\mathcal{L} = \phi\left(1 - \bar{m}\langle P_{\text{hit}}(\sigma', r_0, r)\rangle_{\mathcal{B}(\sigma'_v)}\right) + \kappa \bar{m}, \tag{C12}$$

where $\langle\cdot\rangle_r$ is the mean of over the ball of radius $\sigma'_v$. Minimizing with respect to $\sigma'$ is equivalent to maximizing $\langle P_{\text{hit}}(\sigma', r_0, r)\rangle_r$, which is the probability that a random point in the ball of radius $\sigma'$ and a random point in the ball of radius $\sigma'_v$ are separated by less than $r_0$. This probability is maximized for any $\sigma' \leq \sigma_v - r_0$, where it is equal to $(r_0/\sigma'_v)^d$. Increasing $\sigma'$ beyond $\sigma_v - r_0$ can only lower the probability of recognition. Thus:

$$\min_{\sigma'} \mathcal{L} = \phi + \bar{m}\left(\kappa - \phi \frac{r_0^d}{\sigma_v'^d}\right). \tag{C13}$$

This gives us the condition for the transition from polyclonal to naive, where $\bar{m}^* = 0$. This happens when

$$\text{Polyclonal to naive:} \quad \phi = \kappa \frac{\sigma_v'^d}{r_0^d}. \tag{C14}$$

This condition gives us an exact expression for the location of the transition.

### 3. Exact solution in dimension 1

For $d = 1$, the cost $\mathcal{L}$ in (C4) may be calculated analytically, by using exact expressions of $P_{\text{hit}}(\sigma', r_0, r)$. When $\sigma'_v \leq r_0$, the optimal $\sigma'$ is zero as explained in the main text. When $\sigma'_v > r_0$, we distinguish two cases: $r_0 < \sigma'_v \leq 2r_0$, and $\sigma'_v > 2r_0$.

*a. Case* $r_0 < \sigma'_v \leq 2r_0$. Since we know that the optimal $\sigma'$ is between $\sigma'_v - r_0$ and $\sigma'_v + r_0$, we focus on that range. Then there are two subcases for $\sigma$'.

If $\sigma'_v - r_0 < \sigma' \leq r_0$, there are two contributions to the integral of $\bar{P}_{\text{miss}}$ over the position of the virus $r$. Either $r \leq r_0 - \sigma'$, then all memories recognize the virus, $P_{\text{hit}} = 1$; or $r_0 - \sigma' < r < \sigma'_v < r_0 + \sigma'$, in which case the recognition probability is given by the normalized intersection of two balls at distance $r$ of radii $\sigma'$ and $r_0$,

$$P_{\text{hit}} = \frac{\sigma' + r_0 - r}{2\sigma'}. \tag{C15}$$

Thus we obtain doing the integral over $r$ in (C3):

$$\bar{P}_{\text{miss}}(\sigma', \bar{m}, r_0, \sigma'_v) = \frac{1}{\sigma'_v}\left[(r_0 - \sigma')e^{-\bar{m}} + \int_{r_0-\sigma'}^{\sigma'_v} \exp\left(-\bar{m}\frac{\sigma' + r_0 - r}{2\sigma'}\right) dr\right] \text{ if } \sigma'_v - r_0 < \sigma' \leq r_0. \tag{C16}$$

If $r_0 < \sigma' \leq \sigma'_v + r_0$, there are also two contributions. Either $r \leq \sigma' - r_0$, in which case there is no boundary effect, and the recognition probability is just $P_{\text{hit}} = r_0/\sigma'_v$; or $\sigma' - r_0 < r \leq \sigma'_v \leq \sigma' + r_0$, in which case we have again (C15). Performing the integration in (C3) we obtain:

$$\bar{P}_{\text{miss}}(\sigma', \bar{m}, r_0, \sigma'_v) = \frac{1}{\sigma'_v}\left[\exp\left(-\bar{m}\frac{r_0}{\sigma'}\right)(\sigma' - r_0) + \int_{\sigma'-r_0}^{\sigma'_v} \exp\left(-\bar{m}\frac{\sigma' + r_0 - r}{2\sigma'}\right) dr\right] \text{ if } r_0 < \sigma' \leq \sigma'_v + r_0. \tag{C17}$$

Numerical analysis shows that (C16) admits a minimum as a function of $\sigma'$ in its interval of validity, $\sigma'_v - r_0 < \sigma'^* \leq r_0$, while (C17) is always increasing.

*b. Case* $\sigma'_v > 2r_0$. In this case, there is only a single subcase in the range of interest $\sigma'_v - r_0$ and $\sigma'_v + r_0$. This case is the same as the previous one considered, and the result is given by the same formula (C17). However, for $\sigma'_v > 2r_0$, this expression now admits a minimum $\sigma'_v - r_0 < \sigma'^* \leq \sigma'_v + r_0$.

We recover that in both cases (a and b), in the limit $\bar{m} \to 0$, this minimum is reached at $\sigma'_v - r_0$.

## Appendix D: Population dynamics in sequential immunization

### 1. Clonotype growth and decay as a first-passage problem

We now want to study clonotype proliferation induced by a recall response. We focus on the limit of small mutation rates $\sigma_v \ll r_0$. Within this regime, the system is in the monoclonal phase with $\sigma^* = 0$. We can therefore focus on the case of a single clonotype at position $x = 0$ on the phenotypic space, and ask how successive challenges will modify its size. (Different initial conditions will only change the prefactor in front of the exponential modes in the distribution of first passage times, so the large time behavior of this probability distribution will be the same as discussed below.)

The clonotype has an initial size $n = 1$, and the virus drifts away from $x = 0$ with viral divergence $\sigma_v$. In the general model, cells have probability $\gamma$ to survive from one challenge to the other. Proliferation is taken to be proportional to the cross reactivity radius, $\mu e^{-(r/r_0)^q}$. The population dynamics is thus given by the approximate recursion:

$$n_{t+1} \approx n_t \gamma \left[1 + \mu e^{-(r/r_0)^q}\right], \tag{D1}$$

where we have neglected birth-death noise. We can further simplify this equation to $n_{t+1} = \gamma(1 + \mu\Theta(r - r^*))n_t$, where $r^* = r_0 \ln(\gamma\mu/(1-\gamma))^{1/q}$ is defined as the radius at which the net fold-change factor crosses 1, i.e. when birth is exactly compensated by death. This means that, as long as the virus is within distance $r^*$, the clonotype grows with fold-change factor $\sim \Gamma$. As soon as it reaches $r^*$, and neglecting possible returns below $r^*$ (which happen with probability 1 for $d \leq 2$, but with a frequency that does not affect the overall decay), it will decay with fold-change factor $\sim \gamma$. The problem is thus reduced to determining the first-passage time of the viral antigenic location at radius $r^*$.

We use a continuous approximation corresponding to a slowly evolving strain, $\sigma_v \ll r_0$:

$$a(0) = 0, \qquad da = \frac{\sigma_v}{\sqrt{d}} dW, \tag{D2}$$

where $W$ is a Wiener process. The radius is given by $r(t) = |a(t)|$.

### 2. First passage time in $d = 1$

The distribution of first passage time, $p(t)$, can be computed solving diffusion with a box of size $2r^*$ in $d = 1$ [61]:

$$p(t) = \sum_{n=0}^{+\infty} \frac{(2n+1)\sigma_v^2 \pi}{2r^{*2}} (-1)^n \exp\left(-\frac{(2n+1)^2 \pi^2 \sigma_v^2 t}{8r^{*2}}\right) \tag{D3}$$

The dominant term ($n = 0$) at long times gives an exponential decay:

$$p(t) \approx \frac{\sigma_v^2 \pi}{2r^{*2}} \exp\left(-\frac{\pi^2 \sigma_v^2 t}{8r^{*2}}\right). \tag{D4}$$

### 3. First passage in higher dimensions

We define $f(r, t)$ as the probability density that the virus has not yet reached $r^*$ at time $t$, and is at radius $r$.

This probability density is solution to the diffusion equation with spherical symmetry and absorbing boundary conditions in arbitrary dimension $d > 1$:

$$\frac{\partial f}{\partial t} = \frac{\sigma_v^2}{2d}\left[\frac{\partial^2 f}{\partial r^2} + \frac{d-1}{r}\frac{\partial f}{\partial r}\right], \qquad f(r^*, t) = 0. \tag{D5}$$

Assuming separation of variables, $f(r,t) = T(t)R(r)$, we have:

$$\frac{T'(t)}{T(t)} = \frac{\sigma_v^2}{2d}\frac{R''(r) + \frac{d-1}{r}R'(r)}{R(r)} \equiv -\frac{\sigma_v^2}{2dr^{*2}}\lambda, \tag{D6}$$

where $\lambda$ is to be determined later. This implies $T(t) = Ce^{-\lambda\frac{\sigma_v^2}{2dr^{*2}}t}$ where $C$ is a constant. The radial part $R(r)$ is solution to:

$$R''(r) + \frac{d-1}{r}R'(r) = -\frac{\lambda}{r^{*2}}R(r). \tag{D7}$$

For $d = 1$ this equation reduces to a harmonic equation and we recover the above solution in 1D. Using the change of variable $R(r) = r^{1-d/2}g(r)$ we derive the following equation:

$$r^2 g''(r) + rg'(r) + \left(\lambda\frac{r^2}{r^{*2}} - \left(\frac{d}{2} - 1\right)^2\right)g(r) = 0. \tag{D8}$$

Changing the variable $x = \sqrt{\lambda}r/r^*$, the function $\tilde{g}(x) = g(xr^*/\sqrt{\lambda})$ is solution to the Bessel differential equation of order $d/2-1$. It can therefore be written as a superposition of a Bessel function of the first kind and a Bessel function of the second kind, both of order $d/2-1$. The Bessel function of the second kind having a singularity at $x = 0$, our solution is only given by the Bessel function of the first kind $g(xr^*/\sqrt{\lambda}) = BJ_{d/2-1}(x)$. The radial function $R$ now reads:

$$R(r) = Br^{1-d/2}J_{d/2-1}(\sqrt{\lambda}r/r^*). \tag{D9}$$

The absorbing boundary condition at $r = r^*$ gives us the condition $J_{d/2-1}(\sqrt{\lambda}) = 0$, which has an infinite number of solutions $j_{0,d/2-1}, ..., j_{n,d/2-1}, ...$, so that $\lambda$ can take values

$$\lambda_n = j_{n,d/2-1}^2. \tag{D10}$$

The general solution to (D5) is given as a linear combination of all possible modes, with coefficients $C_n$ determined from boundary conditions and the Dirac delta initial condition, $f(r,0) = \delta(r)$:

$$f(r,t) = \sum_{n=0}^{+\infty} C_n r^{1-d/2} J_{d/2-1}\left(\frac{j_{n,d/2-1}}{r^*}r\right)\exp\left(-\frac{j_{n,d/2-1}^2}{2dr^{*2}}\sigma_v^2 t\right). \tag{D11}$$

The distribution of first passage times asymptotically follows the largest mode of this series, $n = 0$, so that:

$$p(t) \sim \exp\left(-\frac{j_{0,d/2-1}^2}{2dr^{*2}}\sigma_v^2 t\right). \tag{D12}$$

For instance for $d = 2, 3, 4$ we have $j_{0,0} \approx 2.40483$, $j_{0,1/2} = \pi$, $j_{0,1} \approx 3.83171$.

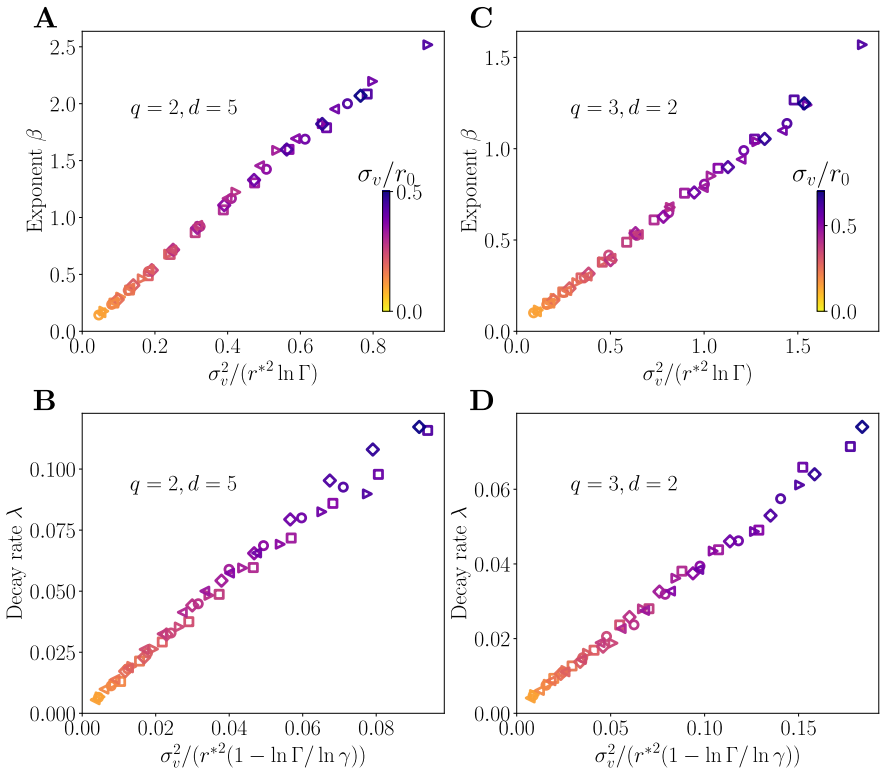


FIG. S2: **Scaling relations for various parameters. A-B.** Power law exponent $\beta$ and lifetime decay rate $\lambda$ in dimension $d = 5$ with a Gaussian cross-reactivity kernel with $q = 2$. **C-D.** Power law exponent $\beta$ and lifetime decay rate $\lambda$ in dimension $d = 2$ with a cross-reactivity kernel with $q = 3$. The different parameters used are $(\gamma = 0.82, \mu = 0.65)$ i.e. $\Gamma = 1.353$ (diamonds), $(\gamma = 0.8, \mu = 0.62)$ i.e. $\Gamma = 1.296$ (squares), $(\gamma = 0.85, \mu = 0.5)$ i.e. $\Gamma = 1.275$ (circles), $(\gamma = 0.87, \mu = 0.4)$ i.e. $\Gamma = 1.21$ (triangles $>$), $(\gamma = 0.9, \mu = 0.35)$ ie. $\Gamma = 1.21$ (triangles $<$). The strategy is optimized with $\phi = 100$, $\kappa = 0.5/(1 - \gamma)$. We used $\alpha = 1$ throughout.