

Critical dynamics in population vaccinating behavior

A. Demetri Pananos¹, Thomas M. Bury¹, Clara Wang², Justin Schonfeld¹, S.P. Mohanty³,
Brendan Nyhan², Marcel Salathé³, Chris T. Bauch^{1*}

1. Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada
2. Department of Government, Dartmouth College, Hanover, New Hampshire
3. Digital Epidemiology Lab, École polytechnique fédérale de Lausanne, Switzerland

* Author for correspondence, cbauch@uwaterloo.ca

Manuscript for submission to *PNAS*, 10 March 2017

Revised manuscript for submission to *PNAS*, 2 November 2017

Major Classification: Biological Sciences

Minor Classification: Ecology

Short title: Critical dynamics in vaccinating behavior

Keywords: socio-ecological systems, machine learning, critical transitions, early warning signals, online social media, vaccine refusal

Abstract

Vaccine refusal can lead to renewed outbreaks of previously eliminated diseases and even delay global eradication. Vaccinating decisions exemplify a complex, coupled system where vaccinating behavior and disease dynamics influence one another. Such systems often exhibit critical phenomena--special dynamics close to a tipping point leading to a new dynamical regime. For instance, critical slowing down (declining rate of recovery from small perturbations) may emerge as a tipping point is approached. Here, we collected and geocoded tweets about measles-mumps-rubella vaccine and classified their sentiment using machine learning algorithms. We also extracted data on measles-related Google searches. We find critical slowing down in the data at the level of California and the United States in the years before and after the 2014-15 Disneyland, California measles outbreak. Critical slowing down starts growing appreciably several years before the Disneyland outbreak as vaccine uptake declines and the population approaches the tipping point. However, due to the adaptive nature of coupled behavior-disease systems, the population responds to the outbreak by moving away from the tipping point, causing "critical speeding up" whereby resilience to perturbations increases. A mathematical model of measles transmission and vaccine sentiment predicts the same qualitative patterns in the neighborhood of a tipping point to greatly reduced vaccine uptake and large epidemics. These results support the hypothesis that population vaccinating behavior near the disease elimination threshold is a critical phenomenon. Developing new analytical tools to detect these patterns in digital social data might help us identify populations at heightened risk of widespread vaccine refusal.

Significance Statement

Complex adaptive systems exhibit characteristic dynamics near tipping points such as critical slowing down--declining resilience to perturbations. We studied Twitter and Google search data about measles from California and the United States before and after the 2014-15 Disneyland, California measles outbreak. We find critical slowing down starting a few years before the outbreak. However, population response to the outbreak causes resilience to increase afterwards. A mathematical model of measles transmission and population vaccine sentiment predicts the same patterns. Crucially, critical slowing down begins long before a system actually reaches a tipping point. Thus, it may be possible to develop analytical tools to detect populations at heightened risk of a future episode of widespread vaccine refusal.

Introduction

In recent decades, vaccine refusal has contributed to the resurgence of measles and pertussis and significantly delayed the global eradication of polio (1, 2). For instance, the 2014-15 measles outbreak in Disneyland, California was preceded by declining kindergarten measles-mumps-rubella (MMR) vaccine coverage in California between 2010 and 2014 (Fig. 1a) (3). Vaccine compliance at school entry fell to 70%-90% in many cases and sometimes even lower in some Los Angeles schools (3). Inadequate vaccine compliance appears to have played a role in the outbreak (4), contributing to a significant peak in California measles case notifications in late 2014 and early 2015 (Fig. 1a) (5). The outbreak garnered significant public interest, causing a large spike in both United States (US)-geocoded tweets regarding measles (Fig. 1b) and Google Internet searches in California for “MMR” and “measles” (Fig. 1c) as reports of cases began to flow in. Amidst the resulting public outcry, the California legislature began taking steps to disallow non-medical exemptions (6–8), although statewide MMR vaccine uptake began to recover before these policy changes went into effect (Fig. 1a) (3).

The changes in vaccinating behavior before and after the Disneyland measles outbreak are consistent with a coupled behavior-disease dynamic in which vaccinating decisions and disease dynamics influence one another in a nonlinear feedback loop. The mathematical modeling of coupled behavior-disease dynamics is growing rapidly (9–12) although relatively little attention has been devoted to critical phenomena in such systems. The theory of critical transitions (tipping points) and their early warning signals may help public health officials anticipate when and where resistance to vaccination might develop and intensify. A critical transition occurs when a complex system shifts abruptly to a strongly contrasting state as an external driver moves the system past a bifurcation point (13, 14). These shifts may exhibit characteristic early warning signals as a consequence of critical slowing down (CSD), in which a declining rate of recovery from small perturbations causes dynamics to become more variable. Critical slowing down can be detected by changes in indicators such the variance, lag-1 autocorrelation, and coefficient of variation in high-resolution time series of state variables (13, 14).

Social norms tend to reinforce currently accepted behavior and thus promote *status quo* practices in populations (15–17). However, individuals also make vaccinating decisions based on the perceived risks of the vaccine and the diseases they prevent (15). Here, we hypothesize that coupled behavior-disease systems exhibit a tipping point arising from interactions between social norms, perceived vaccine risk, and perceived disease risks. Specifically, we investigate the effects of risk perception in terms of the ratio of the magnitude of perceived vaccine risk to the magnitude of perceived risk of disease complications (we will call this “relative vaccine risk” for short). Rising public concern about potential vaccine complications can cause the relative vaccine risk to grow to a tipping point where social norms in support of a *status quo* of high vaccine acceptance can no longer prevent a drop in pro-vaccine sentiment. If the population moves beyond this tipping point, a decline in pro-vaccine sentiment causes fewer people to seek vaccination and herd immunity breaks down, enabling outbreaks of various sizes. But before the tipping point is reached, critical slowing down causes the variance, lag-1 autocorrelation and coefficient of variation of time series of population sentiment toward the vaccine to increase. Importantly, the increase in these three indicators should be noticeable long before any significant change is obvious in the raw time series of population sentiment toward the vaccine. In other words, they provide an early warning signal of a potential tipping point.

However, coupled behavior-disease systems are complex adaptive systems, which introduces an important twist to our hypothesis. The relative vaccine risk is not simply an external driver pushing the system past a tipping point. It also responds to changes in infection prevalence. When an outbreak occurs, the relative vaccine risk drops. Hence, a critical transition can be avoided if the population responds to the small outbreaks that begin to occur near a tipping point (18). We hypothesize that these dynamics could lead to critical slowing down before the outbreak followed by “critical speeding up” (improving resilience to perturbations) *after* the outbreak as the population recedes from the tipping point. Although critical slowing down in a time series of population vaccine sentiment will not necessarily predict whether the population will pull back from the critical transition or go through the transition, it can at least tell us that the population is getting dangerously close to a tipping point.

In this article, we report evidence for critical slowing down in sentiment-classified tweets and in Google searches about measles before the Disneyland measles outbreak, followed by critical speeding up afterwards. These empirical digital signals show patterns that match those

exhibited by a mathematical model of the coupled dynamics of measles transmission and vaccine sentiment that has been previously tested against case notification and vaccine uptake data for measles and pertussis (19–21). Hence, these digital signals could be used as an early warning signal of tipping points in coupled behavior-disease systems.

Results

Model. The mathematical model captures the interplay between disease dynamics, social learning, social norms and perceived risk:

$$\frac{dS}{dt} = \mu(1 - x) - \mu S - \beta SI, \quad [1]$$

$$\frac{dI}{dt} = -\mu I + \beta SI - \gamma I, \quad [2]$$

$$\frac{dx}{dt} = \kappa x(1 - x)(-\omega(t) + I(t) + \delta(2x - 1)), \quad [3]$$

where S is the proportion of susceptible individuals; I is the proportion of infected individuals, x is the proportion of individuals with pro-vaccine sentiment; μ is the per capita birth and death rate, β is the transmission rate, γ is the rate of recovery from infection, κ is the social learning rate, δ is the strength of social norms, and $\omega(t)$ is the relative vaccine risk. We note that Eq. [3] has been rescaled and that the proportion of recovered individuals R is simply $1-S-I$. From Eq. [1], vaccine uptake is given by x and thus all pro-vaccine individuals choose vaccination, while the remainder $1-x$ of anti-vaccinators avoid it. Pro-vaccine sentiment becomes more widespread when infection prevalence $I(t)$ is higher or when vaccine risk $\omega(t)$ is lower. Social norms reinforce whichever sentiment--pro- or anti-vaccine--is more common.

We chose a simple model because critical slowing down only requires that the eigenvalue go to zero at the bifurcation point. This is universal to many types of local bifurcations in both simple and complex models (14). Hence, a broad class of more complicated models should predict the same patterns. (For instance, it is possible to show that including a third category of individuals with neutral sentiment also exhibits critical slowing down). Additional details about model derivation, parameterization and simulation appear in SI Appendix: Section S3.

In the case of fixed vaccine risk, $\omega(t)=\omega$, the model has multiple stable equilibria (19). The equilibrium $(S, I, x)=(0,0,1)$ is of particular interest because it corresponds to a disease-free state with full vaccine uptake that is stable when relative vaccine risk is less than the strength of

social norms ($\omega < \delta$). However, as ω increases past δ the equilibrium is destabilized through a critical transition at $\omega = \delta$ and the population converges to a state of endemic infection and no vaccine uptake (Fig. 2a). At other parameter values, a drop to endemic infection and intermediate vaccine coverage is also possible.

To study critical slowing down, the model was converted to a stochastic model by including an additive Wiener process (SI Appendix: Section S3). When $\omega(t)$ increases linearly until it crosses the tipping point (Fig. 2b), vaccine uptake collapses and an epidemic occurs (Fig. 2c). But before this happens, the variance, lag-1 autocorrelation and coefficient of variation of the time series of pro-vaccine sentiment (x) increase as the critical transition is approached (Fig. 2d). The increase begins long before any significant change is obvious in the raw x time series and hence they provide an early warning signal of the critical transition. We will show later in the Results section that the proportion of individuals with anti-vaccine sentiment ($1-x$) also exhibits critical slowing down.

Approach. In the next subsection, we compare the temporal evolution of the three indicators in digital social data before and after the Disneyland measles outbreak to the model predictions when the relative vaccine risk $\omega(t)$ increases linearly to the tipping point at $\omega = \delta$ and then decreases linearly back to a baseline level (see SI Appendix: Section S3)--this is intended as a first approximation to how critical slowing down (CSD) might occur before the outbreak, followed by critical speeding up after.

We treated CSD in the time series of number of tweets with pro-vaccine (respectively, anti-vaccine) sentiment as a proxy for CSD in the time series of the proportion of individuals with pro-vaccine (respectively, anti-vaccine) sentiment in the general population (x and $1-x$; note that x is also vaccine uptake in the model). This is supported by research showing a correlation between sentiment of tweets on influenza vaccine and actual influenza vaccine uptake (22), and between discussion of individuals' health status in social media and their actual health status (23). We also show that CSD in total tweets of a given sentiment is a good proxy for CSD in population vaccine sentiment and uptake in a broad class of expanded models in which a critical transition in abundance of individuals with pro- or anti-vaccine sentiment drives an observable change in the number of pro- or anti-vaccine tweets in online social media (SI Appendix: Section S6).

We analyzed three empirical data sets. The US GPS data set included measles-related tweets with latitude and longitude coordinates in the US. The much larger California and US Location Field data sets included measles-related tweets from users indicating a California or US location in their user location field. We used a machine learning algorithm to classify tweet sentiment in the Location Field data sets into pro-vaccine, vaccine, or other. The US GPS data set sentiment was classified using Amazon Mechanical Turk (see Methods).

Pro-vaccine tweets. The time series of pro-vaccine tweets shows evidence for CSD in the years before the Disneyland outbreak (Fig. 3). In the California Location Field data set, we observe that the variance (Fig. 3c), lag-1 autocorrelation (AC) (Fig. 3g) and coefficient of variation (Fig. 3k) all increase significantly before the outbreak. The increase in these indicators begins well before the rolling window used for local temporal averaging reaches the time of the outbreak. Hence, the analysis reveals a long-term trend in indicators beginning several years before the outbreak. We interpret this trend as the system's growing variability as the population approaches a critical transition to widespread reductions in vaccine uptake (Fig. 2).

After the outbreak, however, California responds by receding from the critical transition, rather than being pushed past it to a new dynamical regime of endemic infection and significantly reduced vaccine uptake (as occurred for whole cell pertussis vaccination in the United Kingdom, for instance (21)). This is indicated by a decline in all three indicators after the outbreak (Fig. 3 c,g,k), as well as by a reversal of the declining trend in vaccine coverage (Fig. 1c). The system's resilience to perturbations improves as the population recedes from the tipping point.

The decrease in the indicators after the outbreak is also a useful test of whether underlying changes in the total number of Twitter users over the study time window could be driving the observed increase in the indicators before the outbreak. If this were the case, we would not expect to see a decline in the indicators or the number of raw tweets after the outbreak.

The patterns are similar but not as consistent for the data sets from the much larger US population, as expected. Variance increases for both US GPS and US Location Field data sets (Fig. 3a,b), but lag-1 AC increases only for the US GPS data set (Fig. 3e), and the coefficient of variance increases only for the US Location Field data set (Fig. 3j). After the outbreak, the same

indicators in the same data sets decline (Fig. 3a,b,e,j) while the indicator increases in two of the subpanels (Fig. 3i,f).

The mathematical model shows the same general trends, including a stronger signal for variance than for lag-1 AC or coefficient of variation. The three indicators grow and then decline on average in a pattern similar to that observed in the data, as the perceived relative risk $\omega(t)$ approaches and then recedes from the tipping point (Fig. 3d,h,i). The relative magnitude of change in the indicators is also similar in model and data: changes in variance are largest, followed by coefficient of variation, followed in turn by lag-1 AC. In the model, only 66%, 63% and 67% of stochastic realizations exhibit an increase followed by a decrease in the Kendall tau coefficient for variance, lag-1 AC, and coefficient of variation, respectively.

Anti-vaccine tweets. Similar trends are observed for anti-vaccine tweets (Fig. 4), with a surprising exception. As before, the increasing and then decreasing trend in variance is strongest in both model and the three data sets (Fig. 4a-d). However, using Kendall tau values as the criterion, lag-1 AC increases before the outbreak in only one of the three datasets (the US GPS data set, Fig. 4e) and decreases after the outbreak in only two of the data sets (Fig. 4e,g). Trends in lag-1 AC in the model are correspondingly weak, with many stochastic realizations failing to exhibit the increase and decrease (Fig. 4h).

Surprisingly, the coefficient of variation decreases consistently over most of the pre-outbreak time period in all three datasets (Fig. 4i-k). The model also exhibits this inversion (Fig. 4l), with a decrease in the indicator as the tipping point is approached and an increase as the population recedes from it, on average and in 59% of the stochastic realizations (Fig. 4l). Hence the data sets show a post-outbreak decrease as well, and not all pre-outbreak Kendall tau values are negative at the 5% significance level if the time just before the Disneyland outbreak is included. The decline in the coefficient of variation before the tipping point for anti-vaccine but not pro-vaccine sentiment occurs because the statistic divides the standard deviation by the mean. The mean number of non-vaccinators increases from a small value as the tipping point is approached, while the mean number of vaccinators decreases.

Google Trends. Google Trends (GT) is increasingly used in social science and behavioral research (24) and the study of infectious diseases (25, 26). Our search terms did not permit an

analysis of sentiment, but previous research indicates that salient and controversial issues generate higher search volumes (27–29), including a study finding a significant inverse correlation between MMR vaccination coverage and Internet search activity, tweets and Facebook posts (28). If we assume salient and controversial issues are ones on which population opinion is more divided, we can study critical slowing down in the GT Internet search index concerning measles-related searches. These data are also consistent with critical dynamics near a tipping point. The GT data at the national and state levels generally show the same pattern as the Twitter data, with a rise in indicators before the outbreak and a decline afterward (Fig. 5). Trends are stronger at state than national levels, and for “MMR” rather than “measles” searches, which may reflect the greater volume of GT data on “MMR” than “measles” (Fig. 1b).

Sensitivity Analyses. We generated Figs. 3 and 4 using weekly instead of daily bins. For pro-vaccine tweets (SI Appendix Fig. S1), the variance always increases and then decreases, similar to the daily data. Lag-1 AC shows no trend, or tends to decline before the tipping point. However, lag-1 AC measures changes in memory and this is to be expected in a system where memory is short-lived: the life span of a typical online social media news item is less than 24 hours (30), suggesting daily or sub-daily granularity may be required to detect changes in lag-1 AC. The coefficient of variation exhibits a statistically significant increase and decrease before and after the outbreak. Most of these patterns are repeated in the analysis of anti-vaccine tweets using weekly bins (SI Appendix Fig. S2). Results were also qualitatively unchanged when changing the rolling window width used for temporal averaging (SI Appendix Fig. S3-S11).

We analyzed an extended model that includes seasonal variation in the transmission rate and an Erlang-distributed infectious period, both of which are known to influence disease dynamics (31, 32). We found that the indicator trends were unaffected (SI Appendix Fig. S12). Through a probabilistic sensitivity analysis we found that results are qualitatively unchanged across a broad range of parameter values (SI Appendix Fig. S13). To study when happens when the relative vaccine risk responds to infection incidence, we simulated a variant model where $\omega(t) = a + bI(t)$. This variant exhibited growth and decline in the indicators before and after outbreaks, similar to Figs. 3-5 (SI Appendix Fig. S14-16). To rule out that the observed increase and decrease in the indicators can also happen around ordinary (non-critical) outbreaks we

simulated the model at a fixed value of ω far from the critical point. We found that all indicators were flat both before and after non-critical outbreaks (SI Appendix Fig. S17).

Discussion

This article presents evidence that coupled behavior-disease dynamics near the disease elimination threshold is a critical phenomenon. We analyzed tweets and Google searches and showed how the patterns in the empirical data matched those exhibited by a mathematical model of coupled dynamics of measles transmission and vaccine sentiment and uptake. The three indicators--variance, lag-1 autocorrelation, and coefficient of variance--tended to increase before the Disneyland outbreak due to critical slowing down, and then decrease after the outbreak due to critical speeding up (with the unexpected exception of the coefficient of variation in anti-vaccinators where the trend was inverted). Our model predicts the same trends in a population that approaches but then recedes from a tipping point.

The variance indicator showed the most robust trends. However, the coefficient of variation has the advantage that it inherently adjusts for changes in the mean number of tweets, and therefore does not require further processing of the data through computing a residual time series, as required for variance and lag-1 autocorrelation. The lag-1 autocorrelation tests for changes in system memory (13). This indicator often--but not always--showed the expected trends in our data, and trends were not as strong under weekly binning. We speculate this is either because memory is too short-lived in online social media for changes to be detected in data with daily or weekly granularity, or due to the presence of higher-order autoregressive processes that cannot be detected by lag-1 autocorrelation (33, 34).

The Disneyland outbreak was small and the response in population vaccine uptake rapid compared to other episodes of vaccine refusal where populations appear to have crossed a threshold into a regime of endemic infection and significantly reduced population-wide vaccine coverage. This latter scenario occurred for MMR vaccine in England & Wales in the 1990s and 2000s (80% minimum coverage) (21); whole cell pertussis vaccine in England & Wales in the 1970s (30% minimum coverage) (21); and oral polio vaccine in northern Nigeria in 2003-2004 (1). In recent years, measles outbreaks larger than the Disneyland outbreak have occurred in

many under-vaccinated European populations (35). The social media response to the Disneyland outbreak was enormous considering the relatively small size of the outbreak. We speculate this was because the outbreak was the largest in California in many years and it started in a major tourist destination.

A limitation of our model is that it does not account for spatial clustering. This is a key aspect given the presence of clusters of non-vaccinators during the Disneyland measles outbreak (3) and it presents an opportunity for further research given the importance of networks in both infection transmission and strategic interactions (36, 37). The growth of clusters of non-vaccinators is not necessarily a competing hypothesis but rather could represent the spatial manifestation of critical dynamics. Spatially explicit models of behavioral dynamics in related systems develop clusters of individuals with homogeneous opinions as the population starts to “bubble” near a critical phase transition (38). Critical slowing down near a phase transition can manifest in similar ways in both spatial and temporal indicators because the underlying process is similar. Hence, the growing clusters of unvaccinated individuals observed before the Disneyland measles outbreak may signify bubbling near a critical phase transition. This hypothesis could be tested through further research on critical transitions in social networks of Twitter users. We also note that spatio-temporal analysis may take advantage of different and potentially better indicators than purely temporal analysis (39). More research is needed to better understand the informational content of the indicators in spatially structured populations and thereby distinguish qualitatively different outcomes, such as a quick and effective population response versus a protracted period of reduced vaccine coverage and endemic infection. Such analysis could incorporate vaccine uptake data if it has good spatial and temporal resolution (3).

A second limitation is our use of critical slowing down in the number of sentiment-classified tweets as a proxy for critical slowing down in vaccine sentiment and uptake in the general population. This assumption could be relaxed by using more detailed models that include a submodel for online social media activity that accounts for how different users generate differing numbers of tweets and how online social media activity interacts with social processes in the general population.

Our empirical results are largely consistent with our model predictions but cannot definitively establish causality. Future research could evaluate out-of-sample model predictions and consider the relationship between contemporaneous indicators of vaccine sentiment, such as

tweets and search data, and observed vaccine uptake. It would also be valuable to consider other events that might affect sentiment dynamics near tipping points and to evaluate whether the significant population response to the Disneyland outbreak depended on its extensive media coverage.

Still, these results suggest that population vaccinating behavior near the elimination threshold can be characterized as a critical phenomenon near a tipping point in a coupled behavior-disease system. Our findings highlight the value of using digital social data to identify early warning signals of critical dynamics in adaptive behavior-disease systems and socio-ecological systems more generally (18). They also demonstrate the value of using dynamical systems theory in data science. The theory of critical phenomena in complex systems may shed light on other study systems represented in very large social media data sets.

Methods

Twitter data. For the US GPS data set we obtained 27,906 measles-related tweets from March 2, 2011, to October 9, 2016, with GPS coordinates in the US. We used Amazon Mechanical Turk to classify the sentiment of these tweets into 10,926 “pro-vaccine”, 2,136 “anti-vaccine”, and 14,844 “other” categories. A tweet was defined as “pro-vaccine” (respectively, “anti-vaccine”) if the tweet content suggested the tweeter had a positive (respectively, negative) sentiment towards vaccines. This included any information about their feelings or opinions toward vaccines or the diseases they prevent. A tweet was placed in “other” if it was neither pro- nor anti-vaccine, for instance because it was irrelevant, ambiguous, or if the sentiment of the tweeter could not be clearly ascertained. Baseline analysis used daily bins. Additional details appear in SI Appendix: Sections S2 and S5. Over the same time period, 11,685,264 tweets had information in the user location field. To generate the Location Field data sets these tweets were geotagged using a modified version of the Geodict library and classified into “pro-vaccine”, “anti-vaccine” and “other” using a linear support vector machine. The classifier obtained precision scores of 80%, 90%, 79% and recall scores of 83%, 82%, 82% for anti-vaccine, other, and pro-vaccine tweets respectively (F1 scores: 81%, 86%, 80%). The process identified 660,477 anti-vaccine, 883,570 pro-vaccine, and 483,636 other tweets in the US data set, and 101,683 anti-vaccine, 112,741 pro-vaccine, and 59,030 other tweets in the California data set.

Baseline analysis used daily bins. Additional details including references appear in SI Appendix: Section S2 and S5. Data are available in SI Datasets S1-S3.

Google Trends data extraction. We analyzed GT search data for January 2011 to December 2015 using the `gtrendsR` (40) package. Unfortunately, the longest range of day-level query data Google provides is three months, which generates results in the arbitrary units of GT data that are not comparable between searches. (GT returns an estimate of the relative prevalence of searches matching the query for the time period and geography in question when the prevalence of the search term or terms exceeds some unspecified threshold.) As a result, we ran multiple day-level queries for each search (e.g., U.S. “measles,” U.S. “MMR,” California “measles,” California “MMR”) to cover the entire time period and then stacked the resulting data. We then ran a single corresponding week-level query for each search and used this to calculate an adjustment factor (specifically, we multiply each day-level value by the week-level query result divided by the week-level average from the daily data). This adjustment accounts for differences in the relative prevalence of searches over time in the stacked day-level data (41, 42).

CSD indicators. To adjust for long-term changes in the mean number of tweets, we used the residual time series of sentiment-classified tweets for lag-1 AC and variance, generated by subtracting the raw time series from a detrended time series. This is not necessary for the coefficient of variation since it already adjusts for long-term changes in number of tweets. We also removed the Disneyland social media peak (taken as running from January 22 to February 14 based on the US GPS data set) to avoid issues with non-stationarity caused by the Disneyland outlier, and also because our focus is on CSD in the time before and after the outbreak. The methodology of computing indicators for the model was otherwise identical to that for the tweets and GT data. We used the Kendall tau rank correlation to quantify indicator trends (13) although we note that this statistic does not account for the size of increases or decreases over previous time points. Additional details appear in SI Appendix: Section S4.

Acknowledgments

The authors are grateful to Madhur Anand, Feng Fu, and two anonymous reviewers for helpful

comments on the manuscript. This research was funded by an NSERC Discovery Grant and a CFI grant to CTB.

References

1. Jegede AS (2007) What Led to the Nigerian Boycott of the Polio Vaccination Campaign? *PLoS Med* 4(3):e73.
2. Omer SB, Orenstein WA, Koplan JP (2013) Go Big and Go Fast — Vaccine Refusal and Disease Eradication. *N Engl J Med* 368(15):1374–1376.
3. California Department of Public Health: Kindergarten Immunization Levels Available at: <http://www.cdph.ca.gov/programs/immunize/pages/immunizationlevels.aspx> [Accessed February 4, 2016].
4. Majumder MS, Cohn EL, Mekaru SR, Huston JE, Brownstein JS (2015) Substandard vaccination compliance and the 2015 measles outbreak. *JAMA Pediatr* 169(5):494–495.
5. Centers for Disease Control (2015) Morbidity and Mortality Weekly Report (MMWR). *Centers for Disease Control*. Available at: <http://www.cdc.gov/mmwr/> [Accessed November 22, 2016].
6. Whitman E (2015) California Vaccine Bill SB 277: Ban On Personal Exemptions Sparks Counter Movement Despite Recent Measles Outbreak. *International Business Times*. Available at: <http://www.ibtimes.com/california-vaccine-bill-sb-277-ban-personal-exemptions-sparks-counter-movement-1931383> [Accessed November 21, 2016].
7. Siripurapu A (2016) California’s new child vaccination rule takes effect. *Sacramento Bee*. Available at: <http://www.sacbee.com/news/politics-government/capitol-alert/article87023212.html> [Accessed November 21, 2016].
8. Pearlstein J (2016) California’s Pro-Vaccination Law May be Working. *Wired.com*. Available at: <https://www.wired.com/2016/01/californias-pro-vaccination-law-may-be-working/> [Accessed November 21, 2016].
9. Funk S, Salathé M, Jansen VAA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J R Soc Interface* 7(50):1247–1256.
10. Bauch CT, Galvani AP (2013) Epidemiology. Social factors in epidemiology. *Science* 342(6154):47–49.
11. Manfredi P, D’Onofrio A (2013) *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases* (Springer Science & Business Media).
12. Wang Z, Bauch CT, Bhattacharyya S, d’Onofrio A, Manfredi P, Perc M, Perra N, Salathe M, Zhao D (2016) Statistical physics of vaccination. *Phys Rep* 664:1–113.

13. Dakos V, et al. (2012) Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One* 7(7):e41010.
14. Boettiger C, Ross N, Hastings A (2013) Early warning signals: the charted and uncharted territories. *Theor Ecol* 6(3):255–264.
15. Chapman GB, Coups EJ (1999) Predictors of influenza vaccine acceptance among healthy adults. *Prev Med* 29(4):249–262.
16. Streefland P, Chowdhury AM, Ramos-Jimenez P (1999) Patterns of vaccination acceptance. *Soc Sci Med* 49(12):1705–1716.
17. Brunson EK (2013) The impact of social networks on parents' vaccination decisions. *Pediatrics* 131(5):e1397–404.
18. Bauch CT, Sigdel R, Pharaon J, Anand M (2016) Early warning signals of regime shifts in coupled human-environment systems. *Proc Natl Acad Sci U S A* 113(51):14560–14567.
19. Oraby T, Thampi V, Bauch CT (2014) The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc Biol Sci* 281(1780):20133172.
20. Bauch CT (2005) Imitation dynamics predict vaccinating behaviour. *Proc Biol Sci* 272(1573):1669–1675.
21. Bauch CT, Bhattacharyya S (2012) Evolutionary game theory and social learning can determine how vaccine scares unfold. *PLoS Comput Biol* 8(4):e1002452.
22. Salathé M, Khandelwal S (2011) Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol* 7(10):e1002199.
23. Charles-Smith LE, et al. (2015) Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. *PLoS One* 10(10):e0139701.
24. Choi H, Varian H (2012) Predicting the Present with Google Trends. *Econ Rec* 88:2–9.
25. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ (2016) Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proc Natl Acad Sci U S A* 113(24):6689–6694.
26. Milinovich GJ, Williams GM, Clements ACA, Hu W (2014) Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis* 14(2):160–168.
27. Mellon J (2013) Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience. *Journal of Elections, Public Opinion and Parties* 24(1):45–72.
28. Aquino F, et al. (2017) The web and public confidence in MMR vaccination in Italy.

Vaccine 35(35 Pt B):4494–4498.

29. Qi H, Manrique P, Johnson D, Restrepo E, Johnson NF (2016) Open source data reveals connection between online and on-street protest activity. *EPJ Data Sci* 5(1):18.
30. Del Vicario M, et al. (2016) The spreading of misinformation online. *Proc Natl Acad Sci U S A* 113(3):554–559.
31. Earn DJ, Rohani P, Bolker BM, Grenfell BT (2000) A simple model for complex dynamical transitions in epidemics. *Science* 287(5453):667–670.
32. Wearing HJ, Rohani P, Keeling MJ (2005) Appropriate models for the management of infectious diseases. *PLoS Med* 2(7):e174.
33. Ives AR, Dakos V (2012) Detecting dynamical changes in nonlinear time series using locally linear state-space models. *Ecosphere* 3(6):art58.
34. Pace ML, et al. (2017) Reversal of a cyanobacterial bloom in response to early warnings. *Proc Natl Acad Sci U S A* 114(2):352–357.
35. Centers for Disease Control, (CDC) P Increased transmission and outbreaks of measles--European Region, 2011. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/22129994> [Accessed July 6, 2017].
36. Szabó G, Hauert C (2002) Phase transitions and volunteering in spatial public goods games. *Phys Rev Lett* 89(11):118101.
37. Funk S, Gilad E, Watkins C, Jansen VAA (2009) The spread of awareness and its impact on epidemic outbreaks. *Proc Natl Acad Sci U S A* 106(16):6872–6877.
38. Szabó G, Tóke C (1998) Evolutionary prisoner's dilemma game on a square lattice. *Physical Review E* 58(1):69–73.
39. Dakos V, van Nes EH, Donangelo R, Fort H, Scheffer M (2009) Spatial correlation as leading indicator of catastrophic shifts. *Theor Ecol* 3(3):163–174.
40. Perform and Display Google Trends Queries [R package gtrendsR version 1.3.5] Available at: <https://CRAN.R-project.org/package=gtrendsR> [Accessed January 22, 2017].
41. Can we use Daily Internet Search Query Data to Improve Predicting Power of EGARCH Models for Financial Time Series Volatility? (2014). *International Conference on Computer Science and Information Systems (ICISIS'2014) Oct 17-18, 2014 Dubai (UAE)* doi:10.15242/iie.e1014066.
42. Johansen E Creating daily search volume data from weekly and daily data. Available at: <http://erikjohansson.blogspot.com/2014/12/creating-daily-search-volume-data-from.html> [Accessed January 22, 2017].

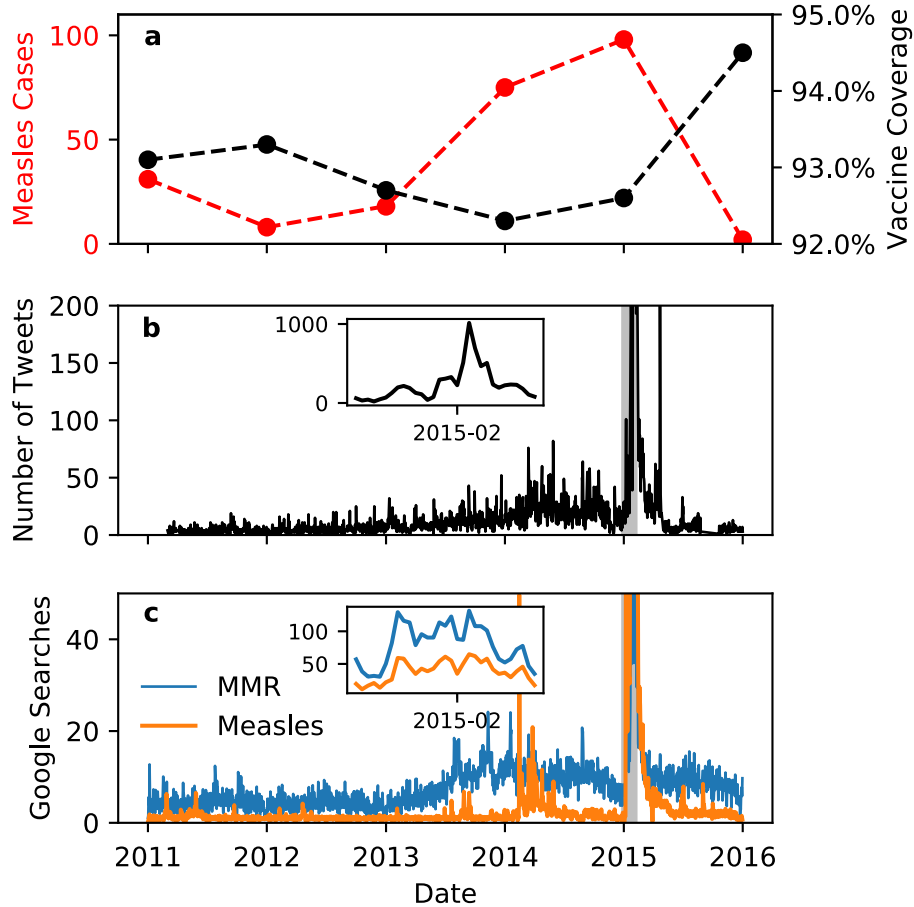
Figures

Fig. 1: Interactions between disease spread, vaccine uptake, and online activity before, during and after the 2014-15 Disneyland, California measles outbreak. (a) Kindergarten MMR vaccine uptake (black; note vertical scale) and measles case notifications in California (red): year in horizontal axis for vaccine uptake corresponds to the ending calendar year of the corresponding academic year (e.g. 2016 means 2015-16 academic year). Case notifications in 2016 go only to Nov. 18. Most 2014 cases occurred at the end of the year. (b) Number of US geocoded tweets for measles-relevant search terms, 2011-16, with a sharp spike in early 2015 corresponding to Disneyland measles outbreak. (c) Google Trends Internet search index for “MMR” (blue) or “measles” (orange) in California, 2011-16, with a sharp spike in early 2015 corresponding to the Disneyland measles outbreak. Shaded region in (b,c) indicates outbreak time period. See SI Appendix: Sections S1 and S2 for details on search terms, data sources, and data extraction.

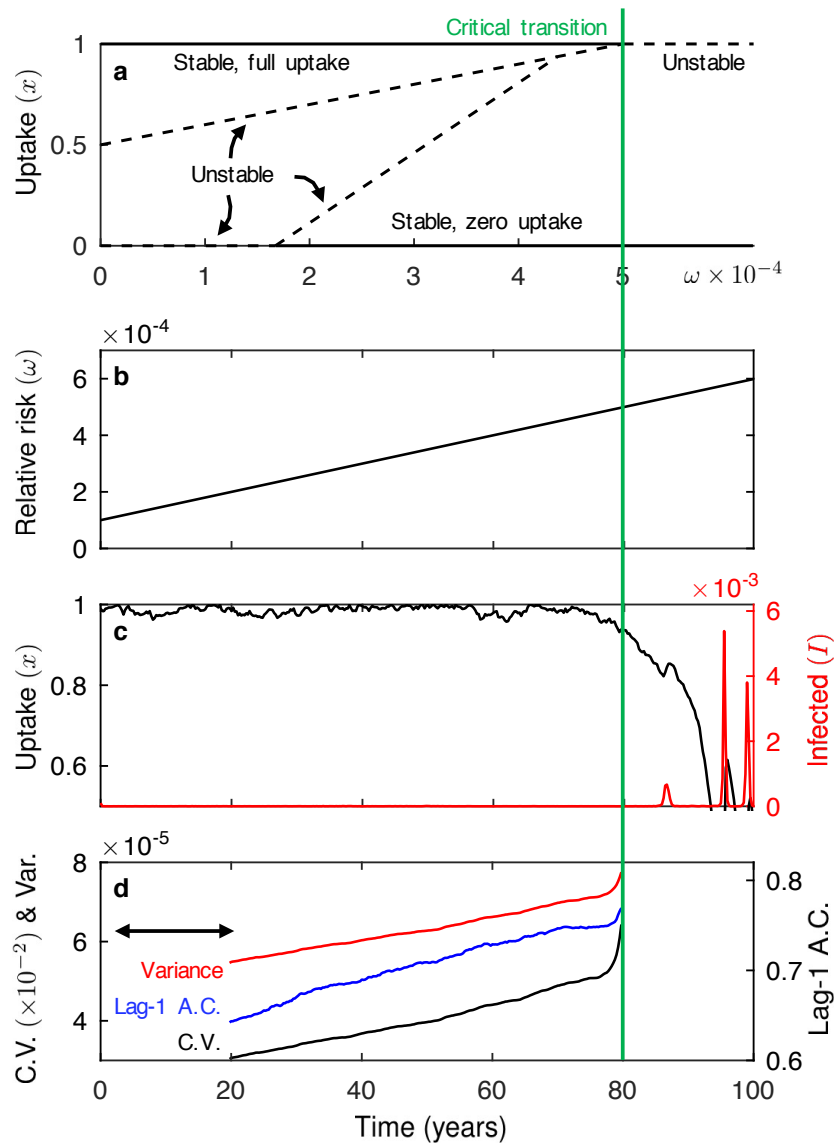


Fig. 2: Coupled behavior-disease model shows early warning signals as perceived risk increases toward a critical transition. Green line indicates location/time of critical transition in all panels. **(a)** Bifurcation diagram of vaccine uptake showing a critical transition from full to zero vaccine uptake when perceived relative risk (ω) exceeds social norm strength (δ) (solid lines are stable branches; dashed are unstable). **(b)** ω (solid line) increasing linearly past critical transition at $\omega=\delta$, **(c)** Vaccine uptake (black) and infection prevalence (red) as ω increases as in **b**. **(d)** Variance (red), lag-1 A.C. (blue), and coefficient of variation (black) for the time series in **c** (mean values at each time point across 500 realizations). Methodological details appear in Methods and SI Appendix: Sections S3, S4.

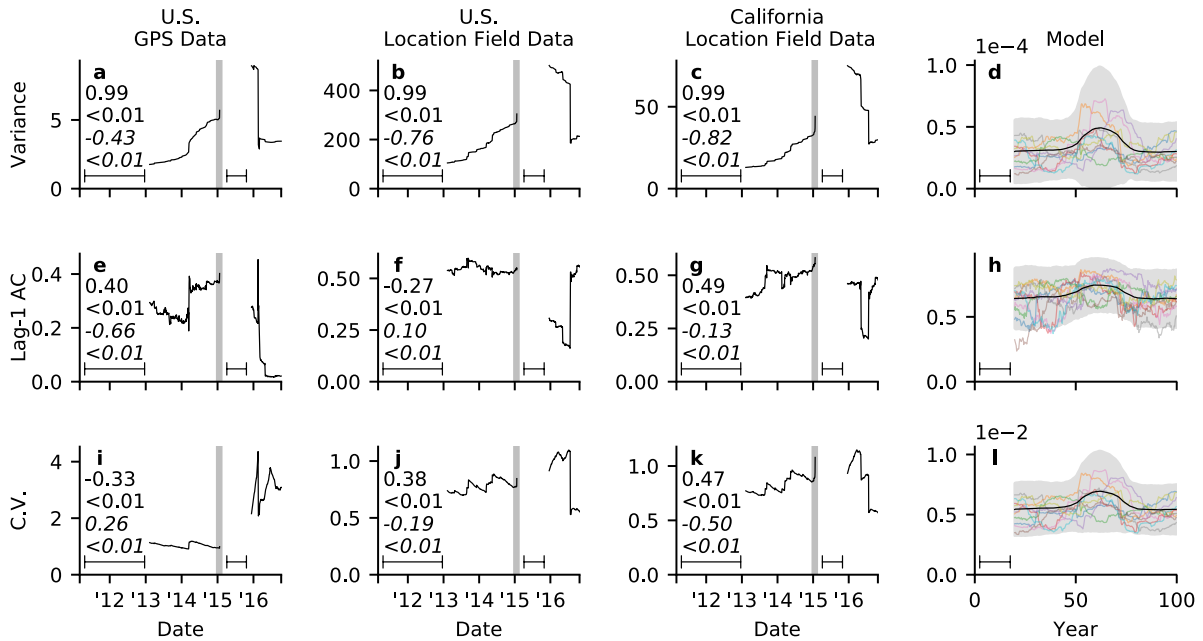


Fig. 3: Critical slowing down pro-vaccine tweets before and after Disneyland measles outbreak. (a-d) Variance, **(e-h)** lag-1 autocorrelation, and **(i-l)** coefficient of variation for **(a,e,i)** US GPS **(b,f,j)** US Location Field, **(c,g,k)** California Location Field data, and **(d,h,l)** model. The residual time series was used for variance and lag-1 AC. Kendall tau rank correlation coefficients are displayed before (regular font) and after (italic) the Disneyland peak with p-values denoted by <. Window width used to compute rolling averages is indicated by line interval. Shaded region indicates outbreak time period. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for details.

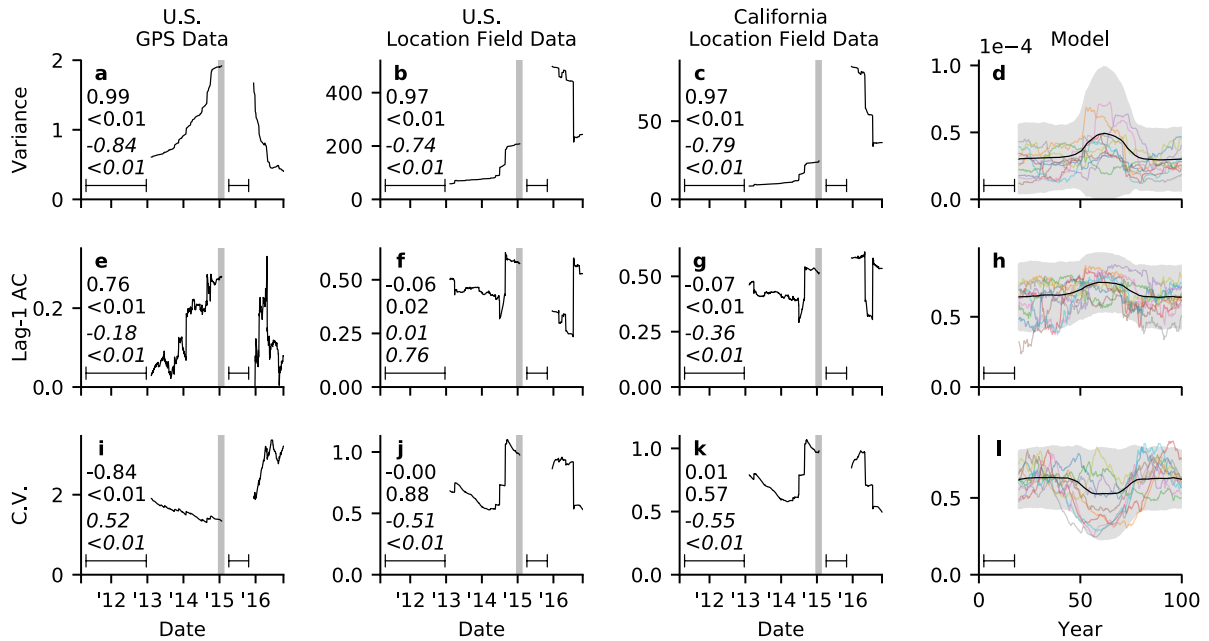


Fig. 4: Critical slowing down in anti-vaccine tweets before and after Disneyland measles outbreak. (a-d) Variance, **(e-h)** lag-1 AC, and **(i-l)** coefficient of variation for **(a,e,i)** US GPS **(b,f,j)** US Location Field, **(c,g,k)** California Location Field data, and **(d,h,l)** model. The residual time series was used for variance and lag-1 AC. Kendall tau rank correlation coefficients are displayed before (regular font) and after (italic) the Disneyland peak with p-values denoted by <. Window width used to compute rolling averages is indicated by line interval. Shaded region indicates outbreak time period. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for details.

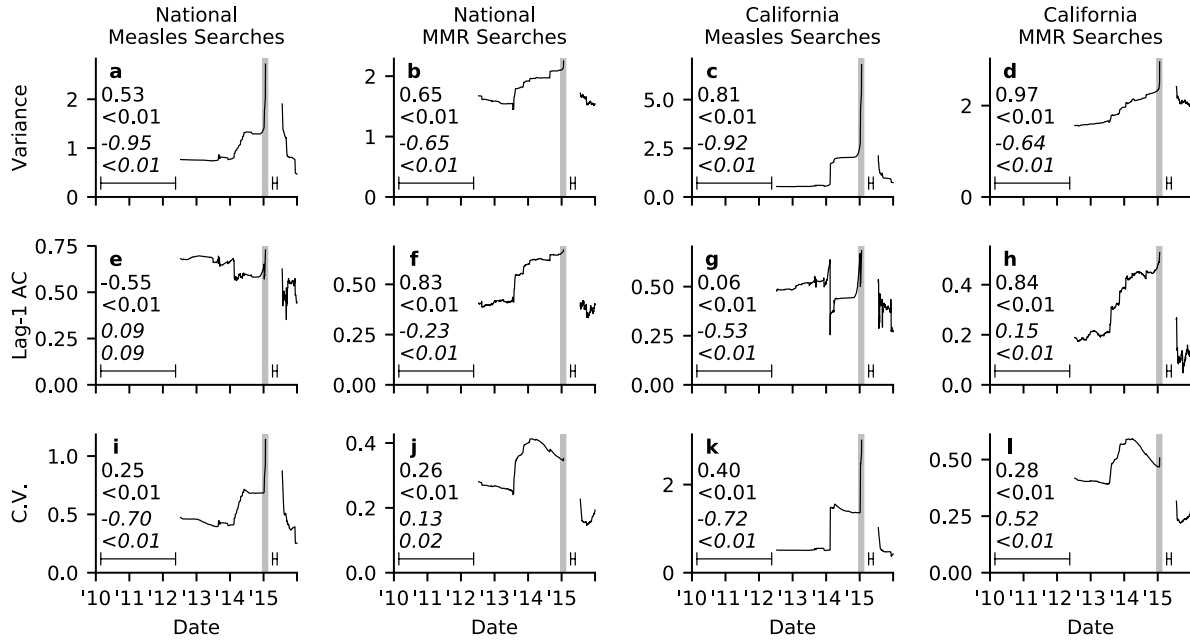


Fig. 5: Critical slowing down in Google Trends search index before and after Disneyland measles outbreak. (a-d) Variance, (e-h) lag-1 AC, and (i-l) coefficient of variation for (a,e,i) US searches for “measles” (b,f,j) US searches for “MMR” (c,g,k) California searches for “measles”, and (d,h,l) California searches for “MMR”. The residual time series was used for variance and lag-1 AC. Kendall tau rank correlation coefficients are displayed before (regular font) and after (italic) the Disneyland peak with p-values denoted by <. Window width used to compute rolling averages is indicated by line interval. Shaded region indicates outbreak time period. See Methods and SI Appendix: Section S4 for details.

Supplementary Information Appendix

Critical dynamics in population vaccinating behavior

A. Demetri Pananos, Thomas M. Bury, Clara Wang, Justin Schonfeld, S.P. Mohanty, Brendan Nyhan, Marcel Salathé, Chris T. Bauch

Section S1. Measles case notifications and measles-mumps-rubella (MMR) vaccine uptake

data. Data on measles case notifications were collected from the California Department of Public Health website (1) and from the CDC Morbidity and Mortality Weekly Report (MMWR) (2). Data on California measles-mumps-rubella (MMR) vaccine uptake in kindergartens at the state and school level were collected from California Department of Public Health reports (1).

Section S2. Extraction of tweets. For the “GPS dataset” (see Methods), the Twitter Application Programming Interface (3) was used to collect all geotagged tweets within the continental United States, given by the bounding box (Latitude 24°, Longitude -126°) and (Latitude 51°, Longitude -66°). The location of the tweets was determined by limiting our search to geotagged tweets (i.e. tweets with precise location information in the metadata). Note that we expect to have close to 100% of the data of interest, as Twitter only limits the results in case a query would result in a dataset that is larger than 1% of the total Twitter data volume. Because our query is rather specific (see below), we are very unlikely to ever run into this limiting problem. A total of 27,906 tweets were collected that were posted from March 2nd, 2011 to October 9, 2016 and contained the following keywords or phrases:

((vaccin* OR immuni*) AND (ingredient* OR risk* OR lies OR disease* OR exemption* OR safe* OR unsafe OR killing* OR conspiracy OR scandal* OR whistleblower* OR pharmaceutical OR CDC OR documentary OR truth* OR theory OR health OR infant* OR baby OR babies OR newborn* OR school* OR aluminum OR death* OR dead OR children OR kid* OR child* OR poison* OR toxic OR mercury OR injur* OR harm* OR brain OR paraly* OR scare* OR fear* OR autism OR IBS OR

autistic or "irritable bowel syndrome")) OR measles OR mmr OR "andrew wakefield"

For the “Location Field dataset” (see Methods), all tweets over the same time period (March 2nd, 2011 to October 9, 2016) and under the same search terms were purchased from GNIP.com. A total of 11,685,264 of these tweets also had content for the location field attached to each tweet. The location field allows twitter account holders to specify their geographic location as unstructured text. One informal study has suggested that 66 percent of twitter account holders put an appropriate place name in the field (4). We extracted place names using a modified version of the Geodict python library (5). The library performs simple extraction by first tokenizing the text, checking the tokens against geographic place databases, and then matching place names against templates such as "CITY, REGION" or "CITY, COUNTRY". A small scale test of the extraction library showed high precision, but low recall. Names that matched multiple localities were resolved in favor of the most populous locale. In the instances where multiple localities were extracted from a single location field all of the localities were considered when determining if the tweet was from an account that identified as being in California or in the United States. From the 11,685,264 tweets with location field information we extracted 689,758 unique location field entries. Of these entries 111,856 contained localities identified as being inside the United States of America, and 11,374 as being inside California. In total, this procedure identified 2,027,683 tweets with a US location, and 273,454 tweets with a California location.

Section S3. Mathematical model parameterization and simulation. We built on an existing deterministic model of coupled behavior-disease dynamics that incorporates social learning, peer pressure, and perceived vaccine and infection risk. The model, which has been tested against case notification and vaccine uptake data for measles and pertussis (6–8), posits that members of the population play either a pro-vaccine or an anti-vaccine strategy. Members engage in social learning by sampling one another at a constant rate; if a member samples someone who is receiving a higher payoff for playing a different strategy, s/he will switch strategies with a probability proportional to the expected gain in payoff (9). The payoff to vaccinate is given by

$e_v = -r_v + \delta_0 x$, where r_v is the perceived risk associated with vaccinating, x is proportion of the population playing the vaccinator strategy, and δ_0 is the strength of injunctive social norms whereby individuals conform to peer pressure. The payoff not to vaccinate is $e_n = -cI(t) + \delta_0(1-x)$, where c is a product of the perceived risk of infection, reporting probability, and a proportionality constant governing the perceived probability of becoming infected, and $I(t)$ is the proportion of the population that is infected. It can be shown that the differential equation governing the dynamics of the proportion of vaccinators, x , equals the difference between these two payoffs multiplied by a term $kx(1-x)$ representing social learning (6). This yields Equation [3] in the main text. Finally, $I(t)$ is obtained from a compartmental SIR model of measles transmission where the vaccine uptake in children is modified as $\mu(1-x)$ (10). This yields Equations [1] and [2] of the main text.

To study critical slowing down, this model was converted into a stochastic model using the relation $dy = F(y)dt + \bar{\sigma}dW$, where y represents S , I , and x , F represents the right-hand sides of Eqs. [1-3], dW is an additive Wiener process and $\bar{\sigma} = [\sigma_s, \sigma_I, \sigma_x]$ is a scaling factor controlling the magnitude of noise for S , I , and x .

Measles parameter values were based on previous literature on compartmental epidemic models and coupled behavior-disease models for paediatric infectious diseases, with $\mu = 0.02$ per year, $\gamma = 365/13/\text{year}$, $R_0 = 16$, and $\beta = R_0(\mu + \gamma)/\text{year}$ (11–14). The strength of social norms, δ , was set to 5×10^{-4} . This value was selected to investigate the worst case scenario in which the critical transition entails a drop from full vaccine coverage to no vaccine coverage. The sampling rate κ was calibrated using the vaccine coverage data in Figure 1. During the time period $\Delta t = 1$ year (from 2015 to 2016), there was an increase in vaccine coverage $\Delta x = 0.02$. Using Euler iteration from the year 2015, where vaccine uptake $x_0 = 0.925$, we have

$$\Delta x = \kappa x_0(1-x_0)[- \omega + I_0 + \delta(2x_0 - 1)]\Delta t.$$

The net payoff for vaccination ΔE (square brackets) is a difficult value to discern for a particular year, however, note that it is on the order of 10^{-4} , and is positive in 2015 since there is a subsequent increase in vaccination coverage. Thus

$$\kappa = \frac{\Delta x}{x_0(1-x_0)\Delta e} \sim \frac{10^{-2}}{10^{-1}10^{-4}} = 1000/\text{yr}$$

Under this parameterization, the endemic branch is unstable and as a result the system will bifurcate from a state of full vaccine coverage to no vaccine coverage.

For the stochastic case used to model critical slowing down, noise levels were set to $\sigma_s=10^{-2}$, $\sigma_I=10^{-5}$, $\sigma_x=10^{-2}$ and initial conditions were set to $(S(0), I(0), x(0)) = (0.01, 10^{-4}, 0.99)$, placing the trajectory in a small neighbourhood of the stable pure vaccinator, disease-free state. The system is studied near the bifurcation point that marks the deterministic transition from high to low vaccine coverage, which occurs at $w(t) = \delta$ (Figure 2a). We investigate two scenarios; one where $w(t)$ increases linearly from a baseline value and crosses the bifurcation point (Figure 2b), the other involves $w(t)$ approaching and then receding from this point (Figures 3 and 4). The linear time evolution of perceived risk $w(t)$ in Figure 2b takes the form

$$w(t) = 1 + t/20,$$

which reaches the bifurcation value at $t = 80$ years. The piecewise relative risk that increases and then decreases (Figures 3 and 4 in main text) take the form

$$\omega(t) = \begin{cases} 1 & 0 \leq t < 40 \\ \frac{2}{5}t - 15 & 40 \leq t < 50 \\ 25 - \frac{2}{5}t & 50 \leq t < 60 \\ 1 & 60 \leq t < 100 \end{cases}$$

which reaches rebounds from the bifurcation point at $t = 50$ years. Additive noise was used in the model to account for case importations.

The coupled stochastic behavior-disease model was simulated in Matlab 2015b using the SDETools Library (15). The SDE was solved using the `sde_euler` function, which implements the Euler-Maruyama Method for numerically solving stochastic differential equations (16). Stochasticity may force the solutions out of physical phase space, however, so solutions were forced to remain in the physical space $[0,1]$ for all model variables. Random seeds were controlled using the `RandSeed` option for the solver to ensure reproducibility. The simulations in Figures 2-4 were run for 100 years with a step size of 5×10^{-4} .

The baseline behaviour-disease model assumes exponential waiting times for the recovery time post infection. However, the infectious period is known to be more constant than this, suggesting the Erlang distribution (which is narrower and more centred about the mean) as a more reliable choice (17). The Erlang distribution is equivalent to a sequence of independent and identically distributed exponential distributions, meaning it can be incorporated into the model by simply introducing multiple ‘recovery’ compartments (18, 19). The number of compartments n , corresponds to the ‘shape parameter’ of the Erlang distribution and can be inferred from appropriate clinical data. We follow a previous study that uses $n = 20$ (17). The modified model reads

$$\begin{aligned}\frac{dx}{dt} &= \kappa x(1-x)(-\omega + I + \delta(2x-1)), \\ \frac{dS}{dt} &= \mu(1-x) - \beta SI - \mu S, \\ \frac{dI_1}{dt} &= \beta SI - (n\gamma + \mu)I_1, \\ \frac{dI_2}{dt} &= n\gamma I_1 - (n\gamma + \mu)I_2, \\ &\dots \\ \frac{dI_n}{dt} &= n\gamma I_{n-1} - (n\gamma + \mu)I_n,\end{aligned}$$

where $I = \sum I_k$.

As a further extension, we include a transmission forcing term that oscillates seasonally. This can be used to account for factors such as changing contact rates among school children (20, 21). We adopt the functional form $\beta(t) = \langle \beta \rangle (1 + \alpha \cos(\frac{2\pi t}{T}))$, with parameters $T = 1$, $\alpha = 0.1$ and $\langle \beta \rangle$ equal to the original transmission rate.

Section S4. Critical slowing down indicators for mathematical model and empirical data.

For the variance and lag-1 autocorrelation indicators, to adjust for non-stationarity caused by long-term changes in the total number of Twitter users, we smoothed the time series for both the stochastic model and the empirical data using a Nadarya Watson estimator with Gaussian kernel at a bandwidth of 10%, selected based on Silverman’s rule of thumb (22). We then subtracted the smoothed time series from the raw time series to generate a detrended (residual) time series.

For the coefficient of variation indicator, it is not necessary to use the residual time series since the indicator is already normalized by the mean. The Kendall tau rank correlation coefficient (23) was used to measure trends in indicators before and after the critical transition.

In the mathematical model, the variance and lag-1 autocorrelation (AC) of the residual time series and the coefficient of variation of the raw time series were computed (24, 25) using a rolling window of width 20 years. Figures 2-4 show the resulting time series of the three indicators averaged over 500 realizations with different random number seeds, as well as two standard deviations over the 500 realizations, for the two types of risk evolution curves $w(t)$. Note that these standard deviations do not correspond to hypothesis testing but rather as a gauge of stochastic variability in the model dynamics. All analyses were conducted in R using the earlywarnings R library (26).

The procedure for detecting critical slowing down in the Twitter and Google Trends datasets was identical to that used in the model simulations except that the Kendall tau rank correlation coefficient was computed separately for pre- and post-outbreak sections, where pre-outbreak was defined as before January 22, 2015 and post-outbreak was defined as after February 14, 2015 (thereby removing the Disneyland online social media peak). The same settings were used for the sensitivity analyses except where otherwise noted. p -values were computed and appear in figure subpanels. Detrending and computation of the Kendall tau rank coefficient were conducted in Python (24), and computation of indicators were conducted in R.

Section S5. Sentiment analysis of tweets. For both GPS and Location Field datasets, the sentiment of each tweet was classified into one of three mutually exclusive categories of pro-vaccine, anti-vaccine, or “other”. A tweet was considered to be pro-vaccine if the content of the tweet suggested the tweeter had a positive sentiment towards vaccines. This could have included information about their actions, feelings, or opinions toward vaccines or the diseases they prevent. For instance, these tweets were classified as pro-vaccine: “Get your children vaccinated people. #WhatGetsMeMad”; “The isolated measles outbreaks could be so much

worse if people just stop vaccinating their kids. People die from this, it's not a game"; "Baby well check and vaccinations today. Poor lil guy but he's happy now!". A tweet was considered to be anti-vaccine if the content of the tweet suggested the tweeter had a negative sentiment towards vaccines. This could have included information about their actions, feelings, or opinions toward vaccines or the diseases they prevent, including actively minimizing or dismissing the dangers of the infection. The following tweets were classified as anti-vaccine, for instance: "I am at loss as to what is happening, the lid is blown on vaccines as a scam, a money maker, harming people more then helping, FOI proves it"; "Who cares if I didn't vaccinate my daughter? It's not your problem it's not your child, it's MY child and my belief", "Did You Know? Vaccinations are Not required to attend school. You are being Lied to. #NoShotsNoSchoolNotTrue". "Other" included any tweet that is neither clearly pro-vaccine or anti-vaccine. These included ambiguous or irrelevant, as well as objective news headlines that suggest the sentiment of the tweeter is neutral or not known. Examples of "other" include "Johnny didn't get measles @midnight #BecauseIDie", "'@HuffingtonPost: Sorry, Disney's new princess Elena probably doesn't count as Latina <http://t.co/gqHYEM6uV2>"#stop the measles outbreak", "I've been exposed to the measles. #dayshift #mybossreadsmytweets (@ San Francisco General Hospital Emergency Room)"; "Reports from the @CDCgov show last year, the nation saw 644 measles cases in 27 states. @wics_abc20 @foxillinois"; "Wow that hpv vaccine will be the death of me". Many tweets included a mention of a news headline. If there was some sensationalist, motivational or emotion-eliciting words in the headline or if the tweeter commentated on the information, or was twisting or misinterpreting facts, it was classified as either pro-vaccine or anti-vaccine according to the sentiment. For instance, this tweet was classified as pro-vaccine: "@thehill: Chris Christie: Parents should have 'measure of choice' on vaccinating children: <http://t.co/E9jRoYdZm2%oU>. No!you ignoramus", while this tweet was classified as anti-vaccine: "Austrian study demonstrates how vaccinated children have more disease, by @SECRETSTOFTHEFED <http://t.co/ZFABxDvmfM>."

For the GPS dataset, Amazon Mechanical Turk (27) was used to classify the sentiment of tweets. The total number of MTurk crowdworkers was 256, and 3 workers were assigned to classify the

sentiment of each tweet. We defined complete agreement as when all the three crowd workers agree on a particular option for a particular tweet; partial agreement as when at least two crowd workers agree on a particular option and just one disagrees; and complete disagreement as when all the three crowd workers submit three different options for a particular tweet. Workers were in complete agreement in 57.9% of tweets, in partial agreement for 39.5% of tweets, and in complete disagreement in 2.6% of tweets. In cases of partial agreement, we assigned the sentiment of the tweet to be the sentiment assigned by the majority of the workers. In the case of complete disagreement, the sentiment of the tweet was assigned to the “other” category. This procedure identified 10,926 pro-vaccine tweets and 2,136 negative tweets. The remaining 14,844 were classified as “other”.

For the Location Field dataset, the number of tweets was considerably larger, therefore machine learning was used to classify the sentiment of the tweets. The tweets were annotated according to sentiment using a linear support vector machine (28) trained on a data set of 75,000 random vaccine related tweets classified using MechTurk using the same methodology that was used for sentiment classification of the GPS dataset (see above). Only tweets where there was unanimous agreement between the three workers who classified each tweet were used. An equivalent number of tweets was used from each sentiment category (positive, negative, and neutral). The classifier had precision scores of [0.80, 0.90, 0.79] and recall scores of [0.83, 0.82, 0.82] for negative, neutral, and positive tweets respectively. The algorithm generated 660,477 negative US tweets; 883,570 positive US tweets; 483,636 US neutral tweets; 101,683 negative California tweets, 112,741 positive California tweets; and 59,030 California neutral tweets.

Section S6. Critical slowing down in an expanded model. To back the argument that critical slowing down (CSD) in online social media may be used as a proxy for CSD in vaccine sentiment amongst the general population, we demonstrate CSD in an expanded model with compartments that could represent social media activity. Consider an additional variable Y that is some measure of online activity with regards to positive vaccine sentiment. For example Y could represent the frequency of pro-vaccine tweets or pro-vaccine users. Assuming that

dynamics in Y are driven by the sentiment of the general population (x) and are proportional to the net payoff of being pro vs. anti-vaccination (ΔE), we propose the generalized dynamics

$$\frac{dY}{dt} = F(x, Y)\Delta E. \quad [S1]$$

The net payoff $\Delta E = -\omega(t) + I + \delta(2x - 1)$ is derived in Section S3 and $F(x, Y)$ is some function that represents the unknown interplay dynamics between online social media Y and the sentiment of the general population x . This evolution equation for Y is then coupled to the original system given in Eqs. [1-3] of the main text. Note that the current setup assumes the impact of behaviour and opinion in the general population influences tweet sentiment (via the x dependence in [S1]), but the influence of twitter sentiment back on general population behaviour is much weaker (and so the original system remains as is). This assumption of unidirectional influence is supported by the observation that only a relatively small proportion of the population uses Twitter to regularly express views on vaccination (see SI Appendix: Section S2 for data on number of US and Canadian tweets).

For some fixed $\omega < \delta$ (close to the tipping point) we may assume that the expanded system sits in some quasi-stationary state (stationary on a fast time-scale whereby evolution of the risk parameter ω can be assumed constant). In accordance with the disease-free equilibrium of the original model, the adjusted stable state is $\vec{z}^* = (0, 0, 1, Y^*)$ for some pro-vaccine tweet count Y^* . This imposes the conditions

$$F(1, Y^*) = 0 \quad [S2]$$

$$\left. \frac{\partial F}{\partial Y} \right|_{(1, Y^*)} < 0 \quad [S3]$$

which ensure stationarity [S2] and stability [S3] of \vec{z}^* . To determine whether CSD occurs in the variable Y near the tipping point $\omega = \delta$, we incorporate stochasticity into the model by adding a Gaussian white noise (GWN) term to each component of the system. The state of the system $\vec{z} = (S, I, x, Y)$ then evolves according to the stochastic differential equation

$$d\vec{z} = \vec{f}(\vec{z})dt + B d\vec{W}(t) \quad [S4]$$

where \vec{f} constitutes the deterministic dynamics given in Eqns. [1-3, S1], $B = \text{diag}(a_1, a_2, a_3, a_4)$ is the diagonal matrix containing the noise amplitudes a_i , and $d\vec{W}(t)$ is a vector of independent

incremental Wiener processes that generate the GWN. For small noise, dynamics about the stable state z^* are well approximated by the linearization

$$d\vec{z} = Jdt + Bd\vec{W}(t) \quad [S5]$$

where J is the Jacobian matrix of f evaluated at z^* . This may readily be computed as

$$J = \begin{pmatrix} -\mu & 0 & -\mu & 0 \\ 0 & -(\gamma + \mu) & 0 & 0 \\ 0 & 0 & -\kappa(\delta - \omega) & 0 \\ 0 & 0 & \beta(\delta - \omega) & -\alpha(\delta - \omega) \end{pmatrix}$$

where

$$\alpha = -\frac{\partial F}{\partial Y}|_{(1,Y^*)}, \quad \beta = \frac{\partial F}{\partial x}|_{(1,Y^*)}.$$

The eigenvalues of J determine the stability of the equilibrium z^* . They are given by $\lambda_1 = -\mu$, $\lambda_2 = -(\gamma + \mu)$, $\lambda_3 = -\kappa(\delta - \omega)$, $\lambda_4 = -\alpha(\delta - \omega)$ and are all negative in the pre-tipping regime ($\omega < \delta$) confirming the stability of z^* . Statistical properties of the fluctuations about z^* may be obtained analytically from S5, given that it takes the form of an Ornstein-Uhlenbeck Process (29). The covariance matrix Σ of the fluctuations satisfies the continuous Lyapunov equation

$$J\Sigma + \Sigma J^T + BB^T = 0$$

which may be solved using elementary matrix techniques (30). Reading off the diagonal elements give us the variances for each state variable. In particular,

$$\text{Var}(Y) = \frac{a_4^2}{2\alpha(\delta - \omega)} + \frac{a_3^2}{2\kappa(\delta - \omega)} \left(\frac{\beta^2}{\alpha(\kappa(\delta - \omega) + \alpha)} \right)$$

which diverges as the tipping point is approached ($\delta \rightarrow \omega$), indicating that the new variable Y does indeed undergo CSD and should provide the relevant early warning signals.

SI Appendix References

1. California Department of Public Health: Kindergarten Immunization Levels Available at: <http://www.cdph.ca.gov/programs/immunize/pages/immunizationlevels.aspx> [Accessed February 4, 2016].
2. Centers for Disease Control (2015) Morbidity and Mortality Weekly Report (MMWR). *Centers for Disease Control*. Available at: <http://www.cdc.gov/mmwr/> [Accessed November 22, 2016].
3. Twitter Application Programming Interface (API) Available at: <https://dev.twitter.com/rest/public> [Accessed July 6, 2017].
4. “Location” Field in Twitter User Profiles (and an interesting fact about Justin Bieber) | PARC blog Available at: <http://blogs.parc.com/2011/01/location-field-in-twitter-user-profiles-and-an-interesting-fact-about-justin-bieber/> [Accessed July 5, 2017].
5. Warden P Geodict: A simple Python library/tool for pulling location information from unstructured text. *GitHub*. Available at: <https://github.com/petewarden/geodict> [Accessed July 5, 2017].
6. Oraby T, Thampi V, Bauch CT (2014) The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc Biol Sci* 281(1780):20133172.
7. Bauch CT, Bhattacharyya S (2012) Evolutionary game theory and social learning can determine how vaccine scares unfold. *PLoS Comput Biol* 8(4):e1002452.
8. Bauch CT (2005) Imitation dynamics predict vaccinating behaviour. *Proc Biol Sci* 272(1573):1669–1675.
9. Helbing D (2010) *Quantitative Sociodynamics* (Springer Verlag).
10. Anderson RM, May RM (1992) *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press).
11. Oraby T, Thampi V, Bauch CT (2014) The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc Biol Sci* 281(1780):20133172.
12. Bauch CT, Bhattacharyya S (2012) Evolutionary game theory and social learning can determine how vaccine scares unfold. *PLoS Comput Biol* 8(4):e1002452.
13. Earn DJ, Rohani P, Bolker BM, Grenfell BT (2000) A simple model for complex dynamical transitions in epidemics. *Science* 287(5453):667–670.
14. Anderson RM, May RM (1992) *Infectious Diseases of Humans: Dynamics and Control*

(Oxford University Press).

15. Horchler A (2011) SDETools. Available at: <https://github.com/horchler/SDETools> [Accessed August 20, 2016].
16. Kloeden PE, Platen E (1992) Higher-order implicit strong numerical schemes for stochastic differential equations. *J Stat Phys* 66(1-2):283–314.
17. Wearing HJ, Rohani P, Keeling MJ (2005) Appropriate models for the management of infectious diseases. *PLoS Med* 2(7):e174.
18. Krylova O, Earn DJD (2013) Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J R Soc Interface* 10(84):20130098.
19. Lloyd AL (2001) Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc Biol Sci* 268(1470):985–993.
20. Bauch CT, Earn DJD (2003) Transients and attractors in epidemics. *Proc Biol Sci* 270(1524):1573–1578.
21. Earn DJ, Rohani P, Bolker BM, Grenfell BT (2000) A simple model for complex dynamical transitions in epidemics. *Science* 287(5453):667–670.
22. Silverman BW (1986) *Density Estimation for Statistics and Data Analysis*.
23. Kendall MG (1938) A New Measure of Rank Correlation. *Biometrika* 30(1/2):81.
24. Knight WR (1966) A Computer Method for Calculating Kendall's Tau with Ungrouped Data. *J Am Stat Assoc* 61(314):436.
25. Malkin R, Robert M (2007) Measurement and Data Analysis for Engineering and Science (Dunn, P.F.; 2004) [Book Review]. *IEEE Eng Med Biol Mag* 26(6):9–11.
26. Dakos V, et al. (2012) Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One* 7(7):e41010.
27. Buhrmester M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Methodological Issues and Strategies in Clinical Research (4th Ed.)*, pp 133–139.
28. Garreta R, Moncecchi G (2013) *Learning scikit-learn: Machine Learning in Python* (Packt Publishing Ltd).
29. Gardiner CW (1985) Handbook of stochastic methods. Available at: <http://dl.icdst.org/pdfs/files/3d1f9d6513c15bdf1eb1e6d16b05e295.pdf>.
30. Negahbani E, Steyn-Ross DA, Steyn-Ross ML, Aguirre LA (2016) Which System Variables Carry Robust Early Signs of Upcoming Phase Transition? An Ecological

Example. *PLoS One* 11(9):e0163003.

SI Appendix Figures

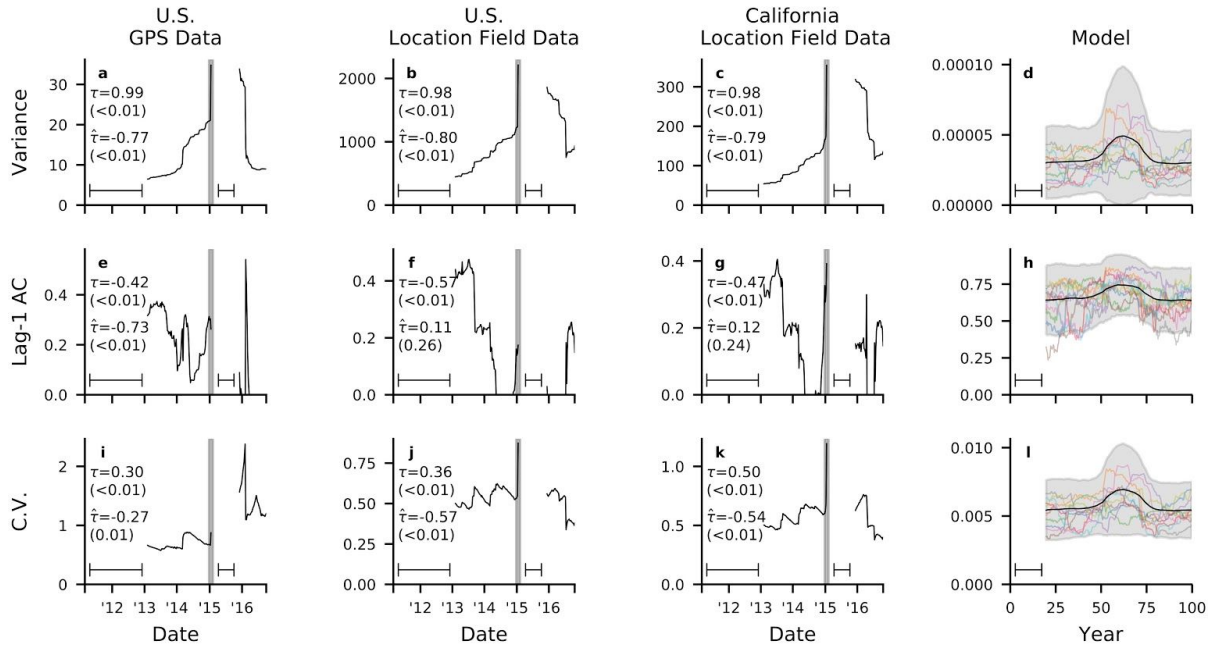


Figure S1: Critical slowing down in time series of pro-vaccine tweets before and after the Disneyland measles outbreak, weekly bins for data. (a-d) Variance, (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation (C.V.) indicators for (a,e,i) US GPS (b,f,j) US Location Field, (c,g,k) California Location Field data, and (d,h,l) mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

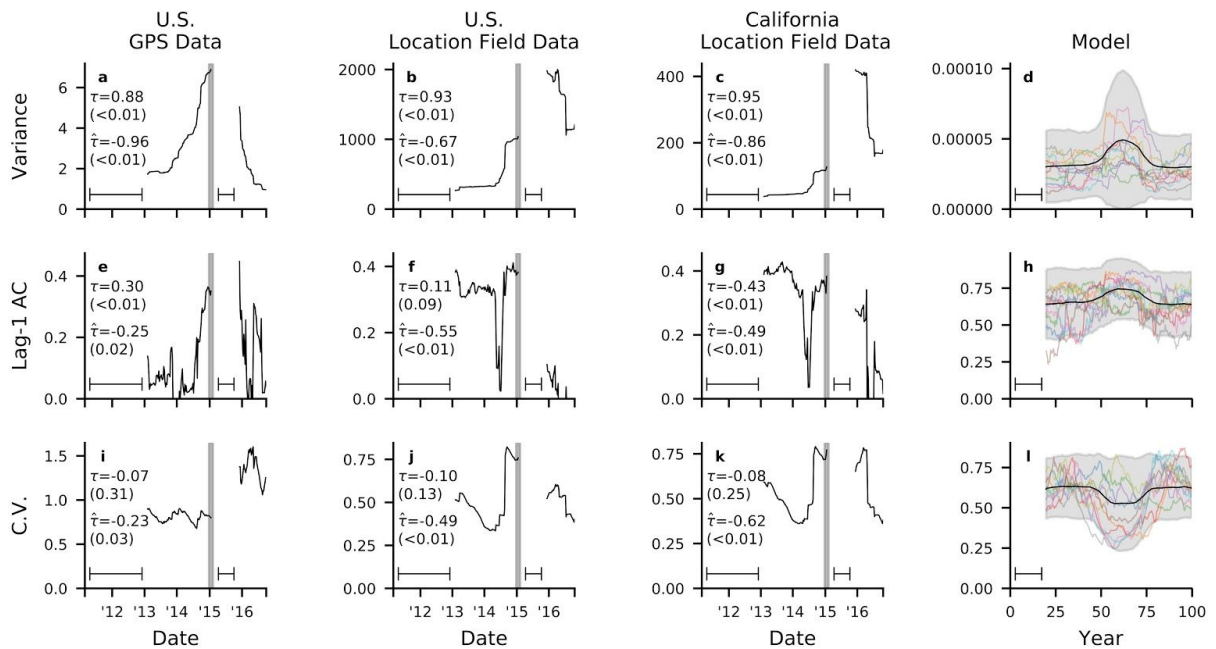


Figure S2: Critical slowing down in time series of anti-vaccine tweets before and after the Disneyland measles outbreak, weekly bins for data. (a-d) Variance, (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation (C.V.) indicators for (a,e,i) US GPS (b,f,j) US Location Field, (c,g,k) California Location Field data, and (d,h,l) mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

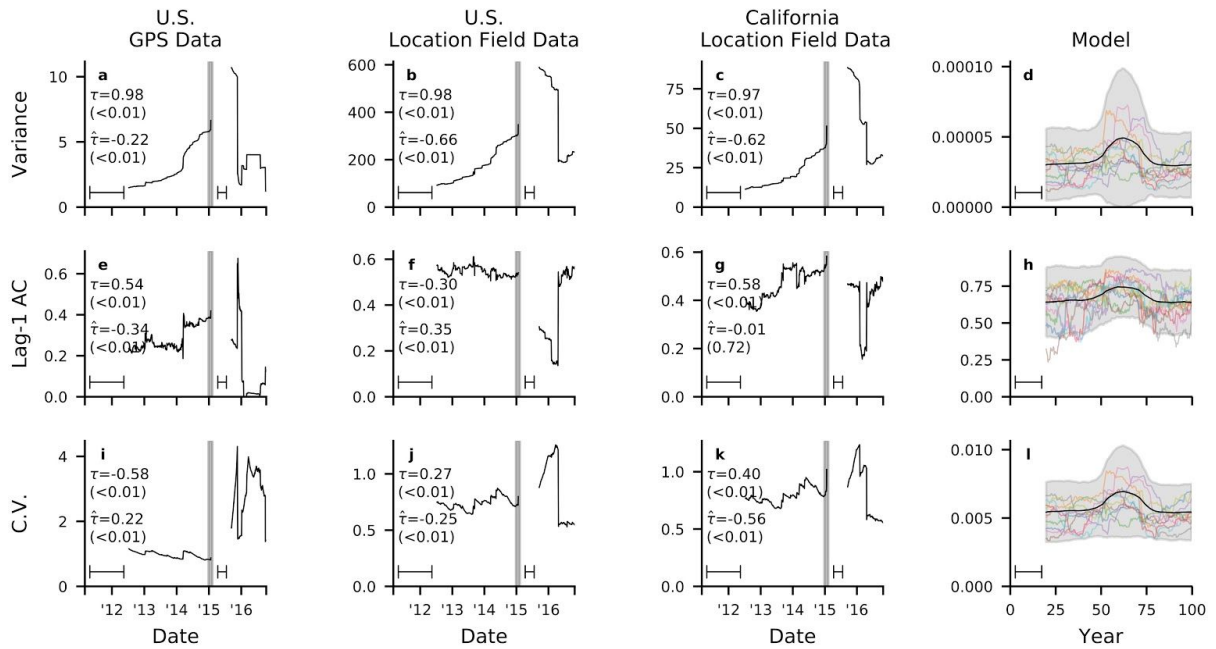


Figure S3: Critical slowing down in time series of pro-vaccine tweets before and after the Disneyland measles outbreak, rolling window width is 35% the width of the dataset. (a-d) Variance, (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation (C.V.) indicators for (a,e,i) US GPS (b,f,j) US Location Field, (c,g,k) California Location Field data, and (d,h,l) mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

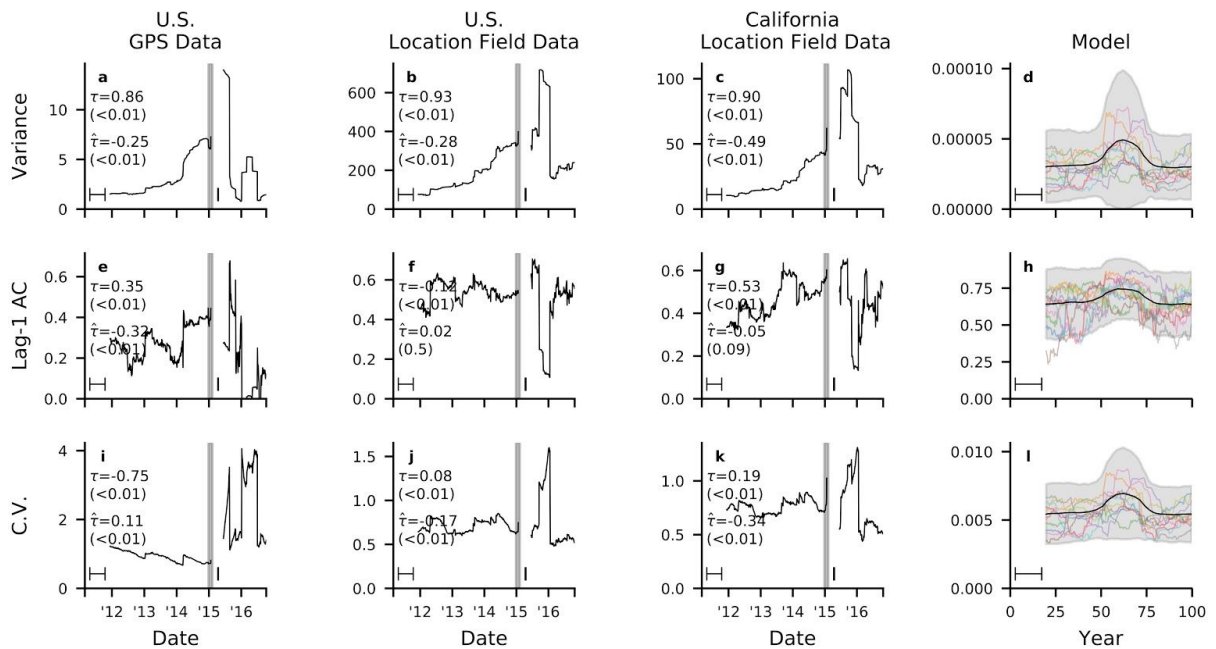


Figure S4: Critical slowing down in time series of pro-vaccine tweets before and after the Disneyland measles outbreak, rolling window width is 20% the width of the dataset. (a-d) Variance, (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation (C.V.) indicators for (a,e,i) US GPS (b,f,j) US Location Field, (c,g,k) California Location Field data, and (d,h,l) mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

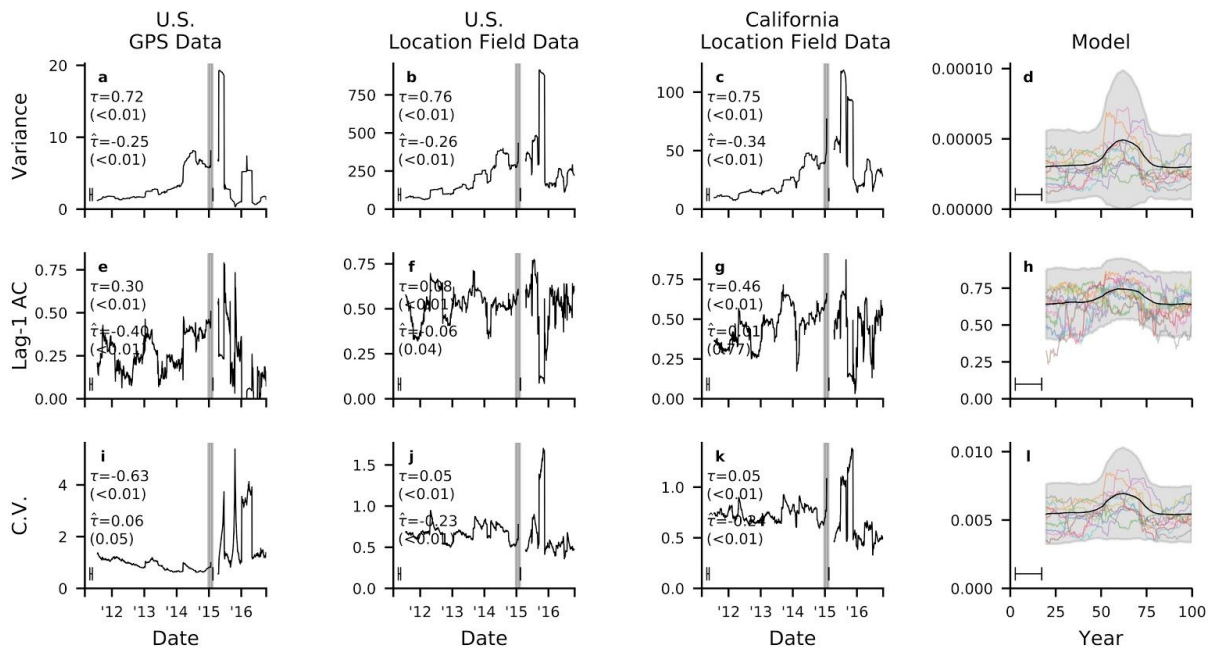


Figure S5: Critical slowing down in time series of pro-vaccine tweets before and after the Disneyland measles outbreak, rolling window width is 10% the width of the dataset. (a-d) Variance, **(e-h)** lag-1 autocorrelation, and **(i-l)** coefficient of variation (C.V.) indicators for **(a,e,i)** US GPS **(b,f,j)** US Location Field, **(c,g,k)** California Location Field data, and **(d,h,l)** mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

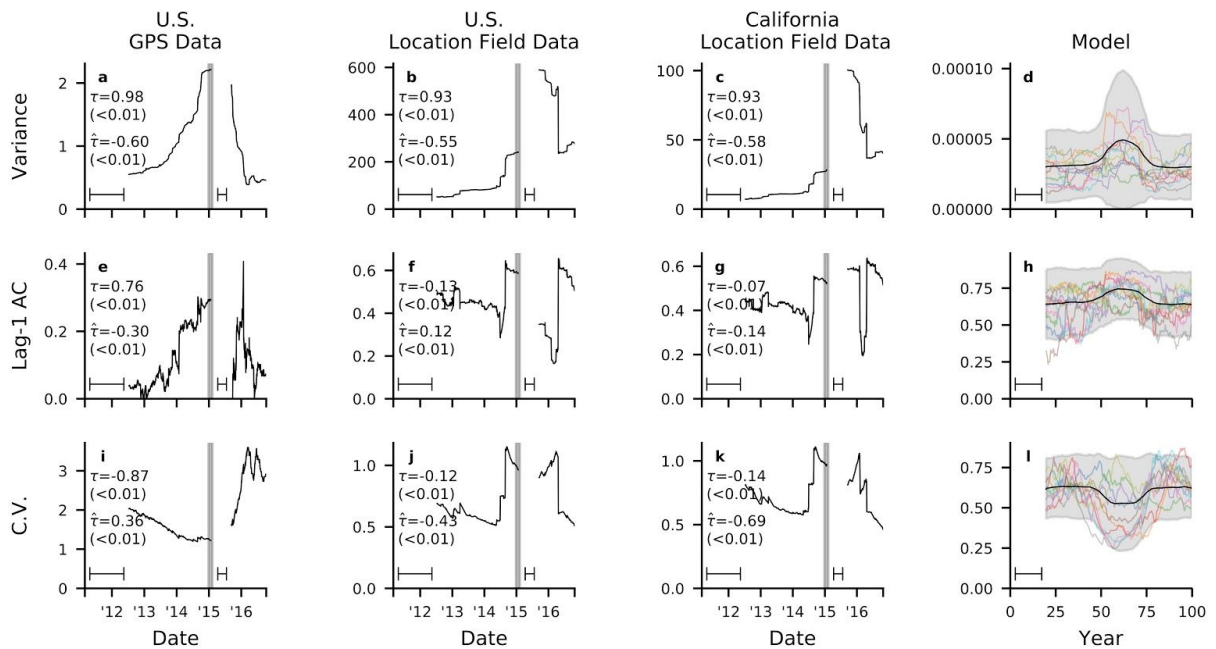


Figure S6: Critical slowing down in time series of anti-vaccine tweets before and after the Disneyland measles outbreak, rolling window width is 35% the width of the dataset. (a-d) Variance, **(e-h)** lag-1 autocorrelation, and **(i-l)** coefficient of variation (C.V.) indicators for **(a,e,i)** US GPS **(b,f,j)** US Location Field, **(c,g,k)** California Location Field data, and **(d,h,l)** mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

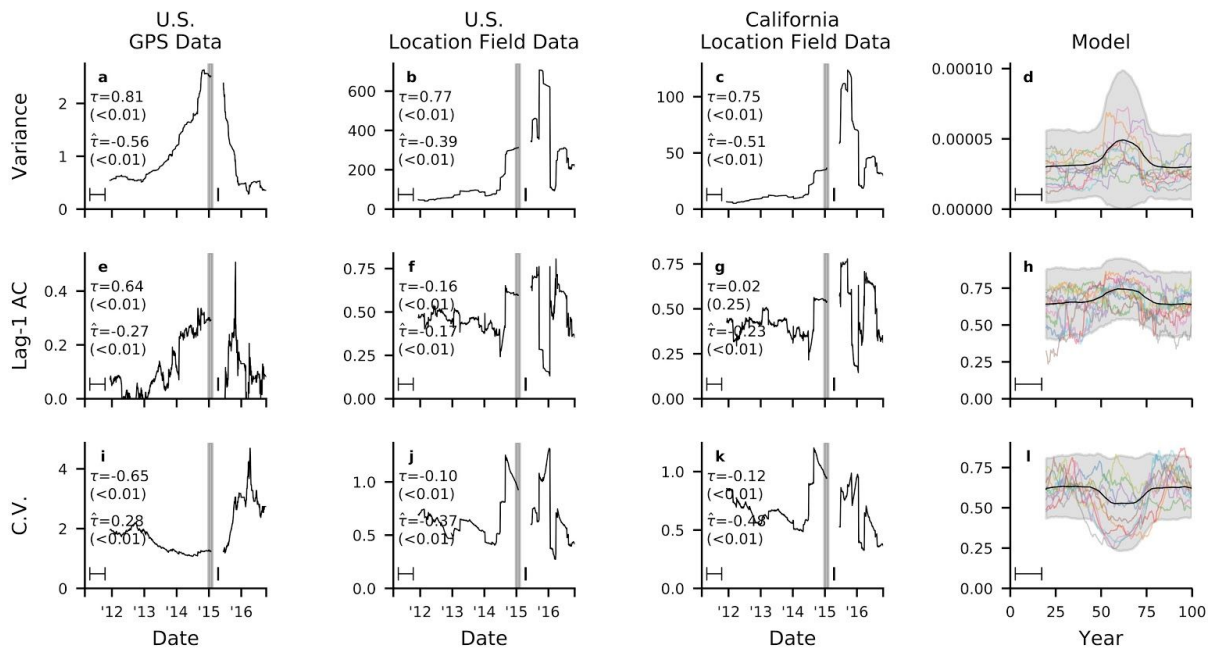


Figure S7: Critical slowing down in time series of anti-vaccine tweets before and after the Disneyland measles outbreak, rolling window width is 20% the width of the dataset. (a-d) Variance, **(e-h)** lag-1 autocorrelation, and **(i-l)** coefficient of variation (C.V.) indicators for **(a,e,i)** US GPS **(b,f,j)** US Location Field, **(c,g,k)** California Location Field data, and **(d,h,l)** mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

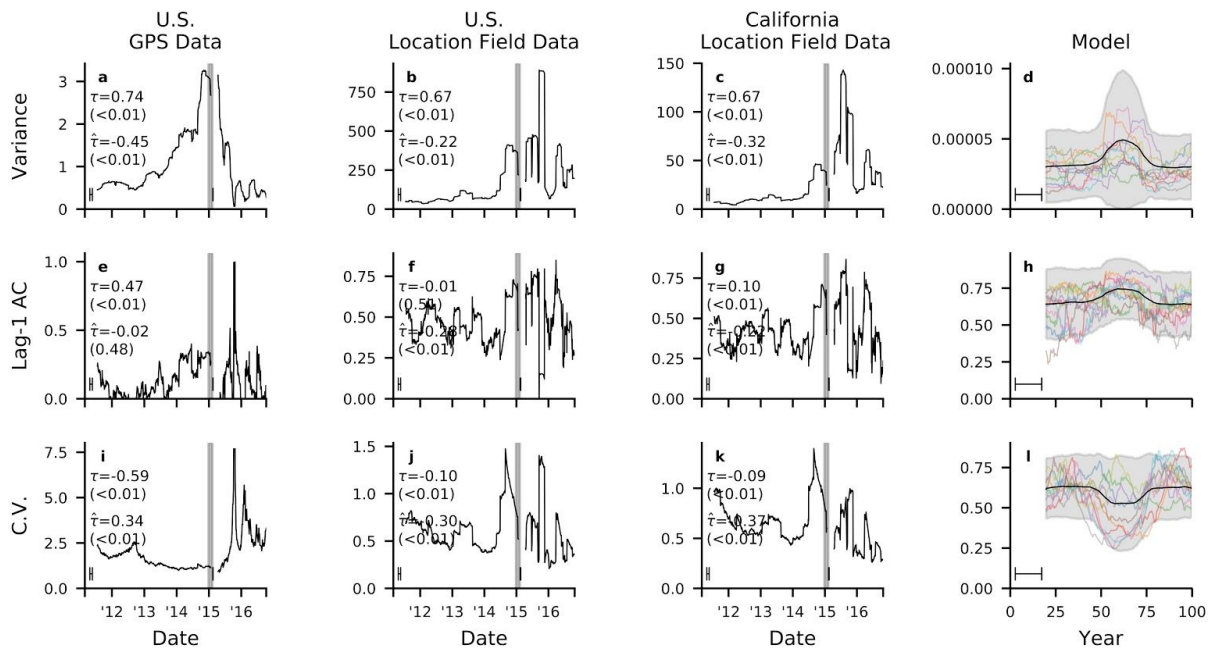


Figure S8: Critical slowing down in time series of anti-vaccine tweets before and after the Disneyland measles outbreak, rolling window width is 10% the width of the dataset. (a-d) Variance, **(e-h)** lag-1 autocorrelation, and **(i-l)** coefficient of variation (C.V.) indicators for **(a,e,i)** US GPS **(b,f,j)** US Location Field, **(c,g,k)** California Location Field data, and **(d,h,l)** mathematical model. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region in data panels indicates time period of the outbreak. Model panels show indicators averaged across 500 stochastic model realizations (black), two standard deviations of the 500 realizations (shaded), and 10 example realizations (colored lines). See Methods and SI Appendix: Sections S3, S4 and S5 for further details.

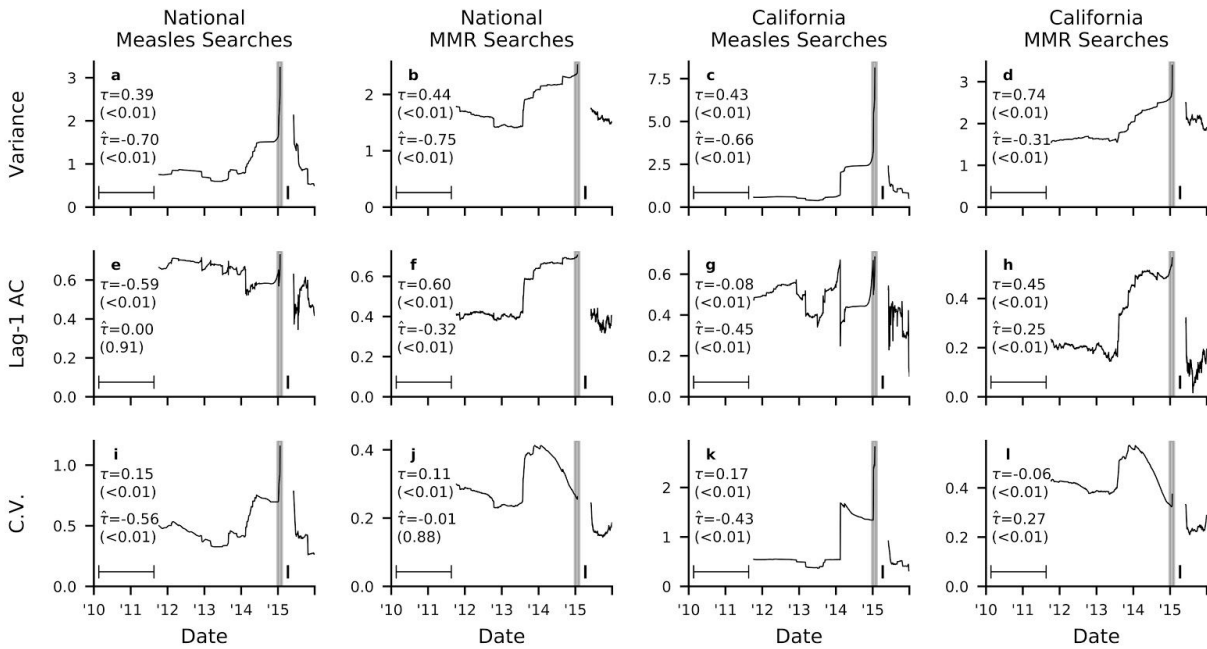


Figure S9: Critical slowing down in Google Trends search index before and after the Disneyland measles outbreak, rolling window with is 35% the width of the dataset. (a-d) Variance, (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation (C.V.) indicators for (a,e,i) US searches for “measles” (b,f,j) US searches for “MMR” (c,g,k) California searches for “measles”, and (d,h,l) California searches for “MMR”. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region indicates time period of the outbreak. See Methods and SI Appendix: Section S4 for further details.

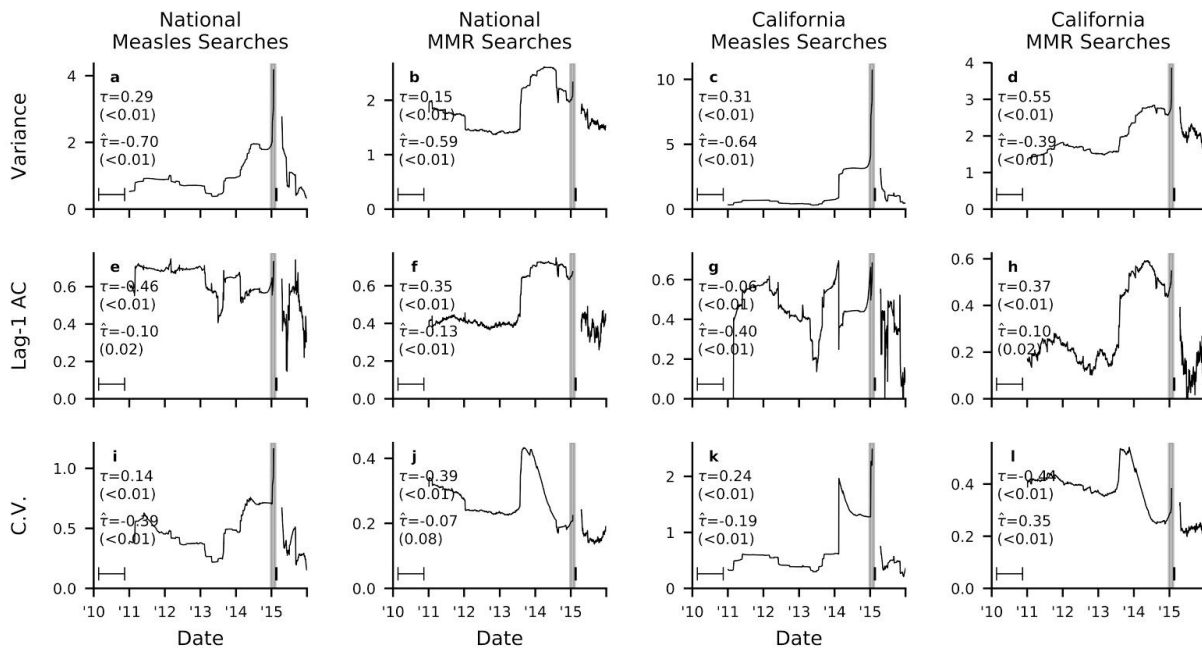


Figure S10: Critical slowing down in Google Trends search index before and after the Disneyland measles outbreak, rolling window with is 20% the width of the dataset. (a-d) Variance, (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation (C.V.) indicators for (a,e,i) US searches for “measles” (b,f,j) US searches for “MMR” (c,g,k) California searches for “measles”, and (d,h,l) California searches for “MMR”. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region indicates time period of the outbreak. See Methods and SI Appendix: Section S4 for further details.

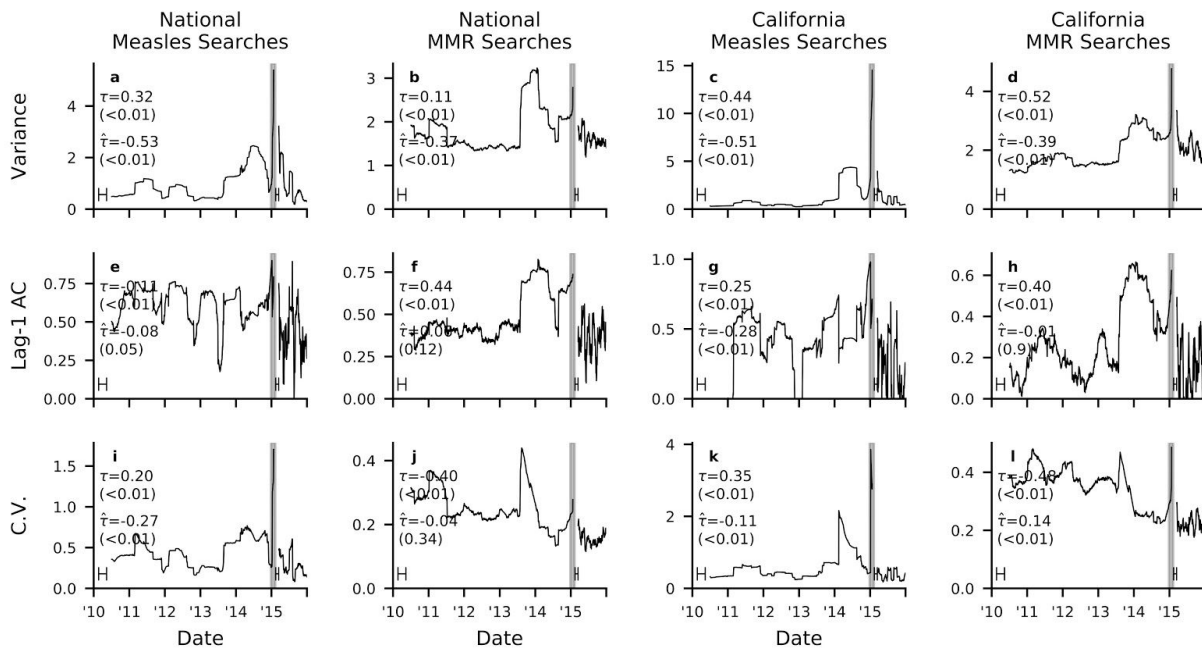


Figure S11: Critical slowing down in Google Trends search index before and after the Disneyland measles outbreak, rolling window with is 10% the width of the dataset. (a-d) Variance, (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation (C.V.) indicators for (a,e,i) US searches for “measles” (b,f,j) US searches for “MMR” (c,g,k) California searches for “measles”, and (d,h,l) California searches for “MMR”. Also displayed are Kendall tau rank correlation coefficients before (τ) and after ($\hat{\tau}$) the Disneyland online peak, with p-values in parentheses. Window width used to compute rolling averages is indicated by line intervals. Shaded region indicates time period of the outbreak. See Methods and SI Appendix: Section S4 for further details.

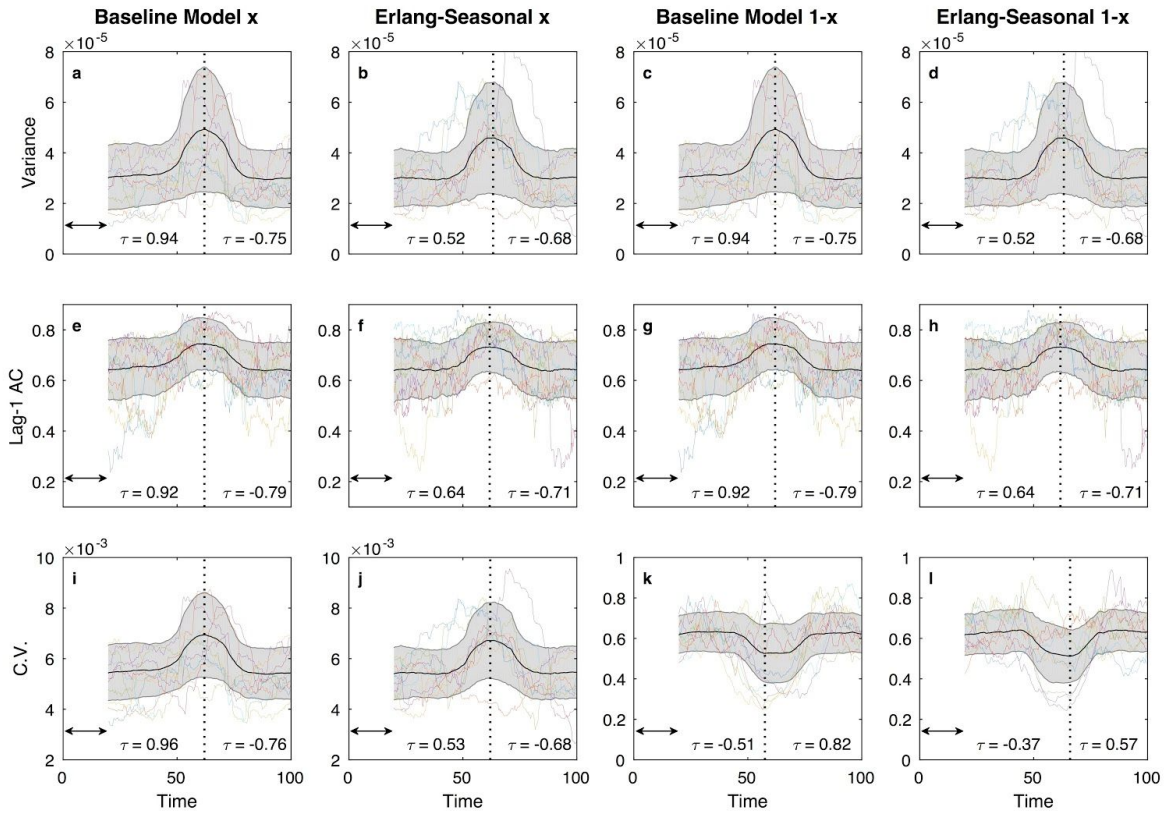


Figure S12: Predicted indicator trends for baseline model compared to extended model with seasonal variation in transmission rate and Erlang-distributed infectious period.

Baseline mathematical model predictions for (a,e,i) x time series and (c,g,k) $1-x$ time series, and extended model predictions for (b,f,j) x time series and (d,h,l) $1-x$ time series. Indicators shown are (a-d) variance; (e-h) lag-1 autocorrelation, and (i-l) coefficient of variation. Panels show indicators averaged across 500 stochastic model realizations (black), one standard deviation of the 500 realizations (shaded), and 10 example realizations (colored lines). Perceived risks increases and decreases as in Figures 3 and 4; dashed line shows time of maximum perceived risk. See Methods and SI Appendix: Sections S3 and S4 for further details.

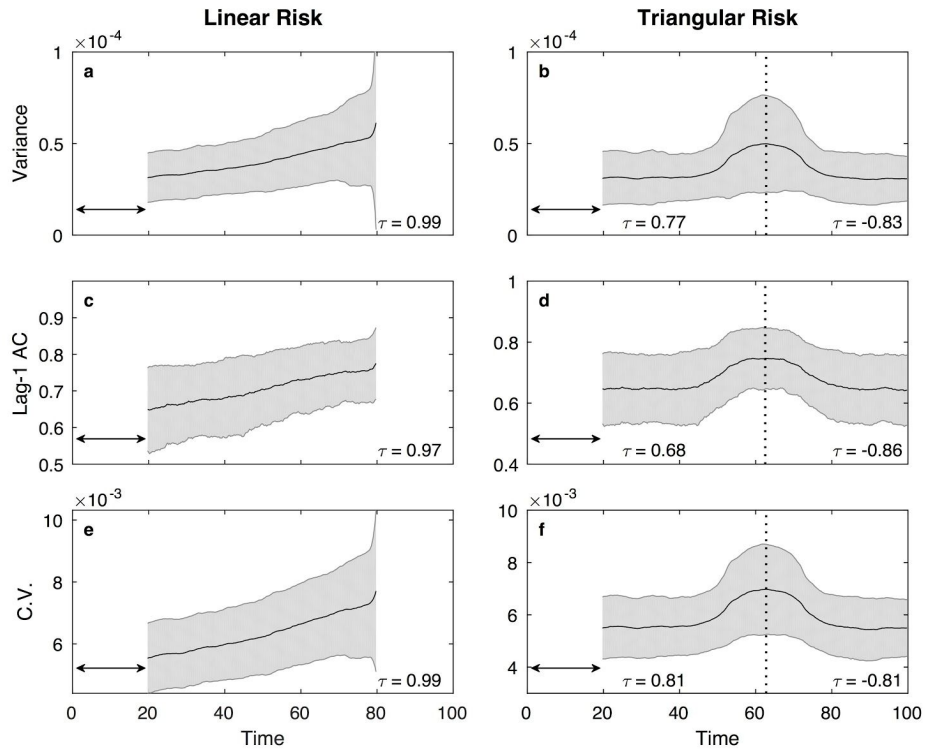


Figure S13: Sensitivity analysis on model parameter values for all three indicators.

(a,b) Variance; (c,d) lag-1 autocorrelation, and (e,f) coefficient of variance of the x time series at a range of possible parameter values when (a,c,e) perceived risk ω increases linearly beyond $\omega = \square$, as in Figure 2 and (b,d,f) perceived risk ω increases linearly to $\omega = \square$ and then declines, as in Figures 3 and 4. Black line shows mean value of 500 stochastic realizations; dotted vertical line shows location of maximal perceived risk; shaded region shows one standard deviation of the 500 stochastic realizations using parameter values sampled from triangular distributions. Epidemiological parameters, as well as the sampling rate k , were drawn from a triangular distribution ranging from half to three halves of the baseline value. The time of onset of the risk evolution curve was also sampled from a triangular distribution. For the linear case, the onset was drawn from a triangular distribution ranging from 8 to 12 years. Likewise, the time of onset of the triangular curve was drawn from a triangular distribution ranging from 20 to 40 years. The triangular curve remained symmetric about the $t = 50$ line, attaining a value of δ when $t = 50$. Double-headed arrow denotes length of rolling window used for computing lag-1 AC. Further details appear in SI Appendix: Section S3 and S4.

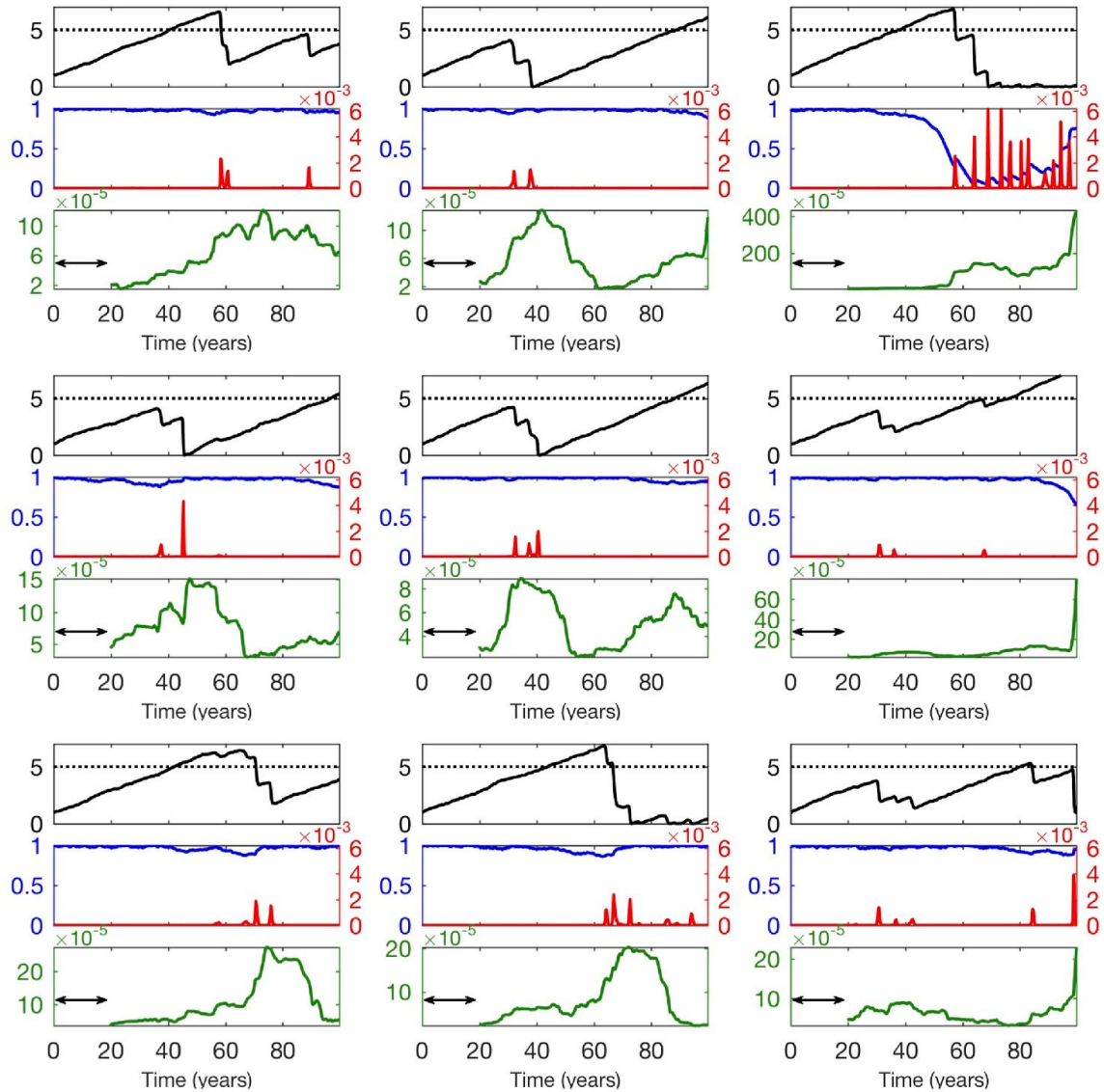


Figure S14: Adaptive dynamics in extended model where perceived risk obeys $d\omega/dt = a + bI(t)$, in nine model realizations, variance indicator. Each set of three subpanels shows the results of a single realization. In each set, top subpanel shows time evolution of $\omega(t)$ (solid line) and location of critical transition (dashed line); middle subpanel shows vaccine uptake (blue) and infection prevalence (red); bottom subpanel shows variance (green) of x time series computed using a rolling window of width indicated by arrows. See SI Appendix: Section S3 and S4 for details on model simulation and computation of variance. $a=1 \times 10^{-4}$, $b=-0.2$.

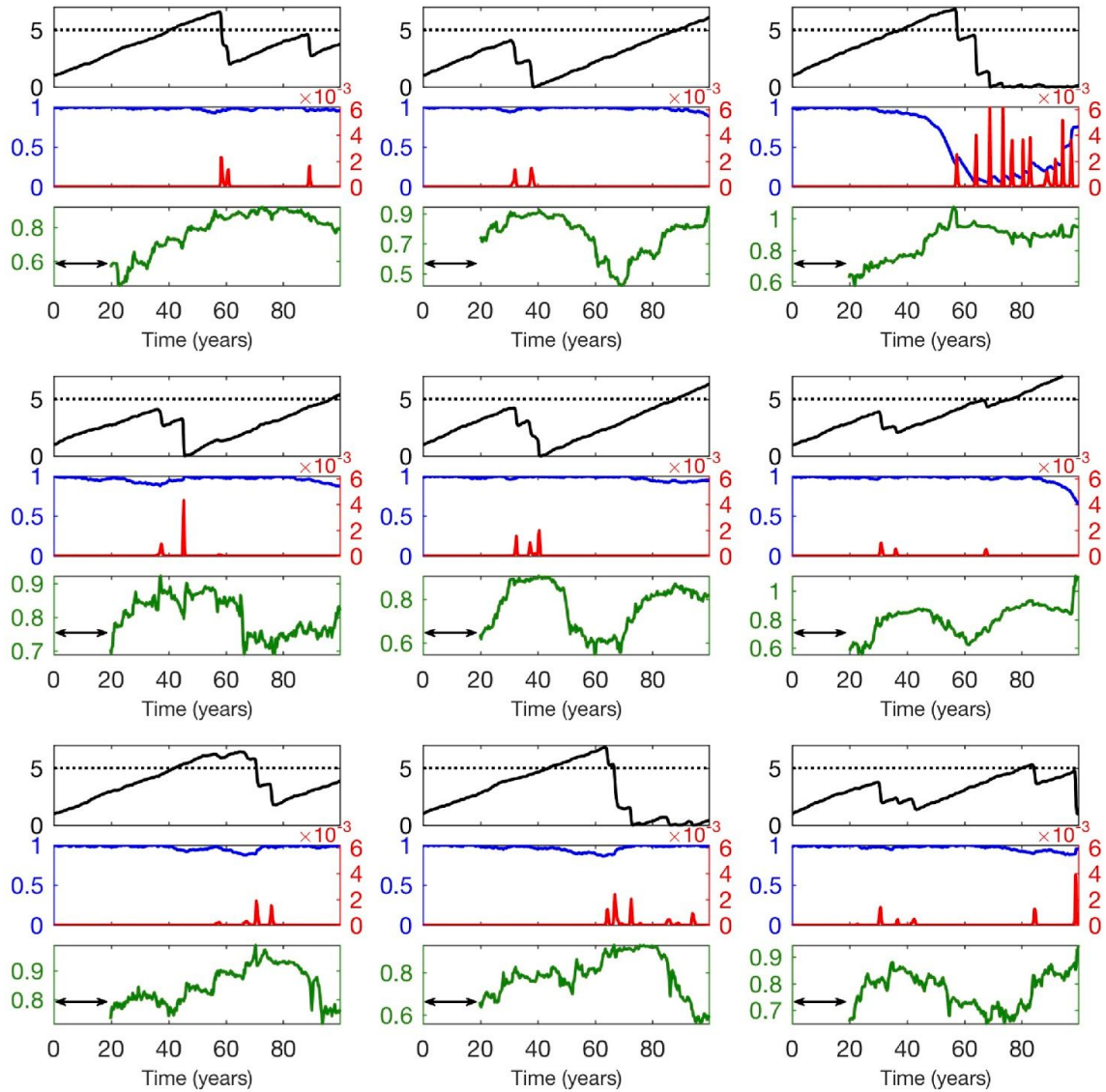


Figure S15: Adaptive dynamics in extended model where perceived risk obeys $d\omega/dt = a + bI(t)$, in nine model realizations, lag-1 autocorrelation indicator. Each set of three subpanels shows the results of a single realization. In each set, top subpanel shows time evolution of $\omega(t)$ (solid line) and location of critical transition (dashed line); middle subpanel shows vaccine uptake (blue) and infection prevalence (red); bottom subpanel shows lag-1 autocorrelation (green) of x time series computed using a rolling window of width indicated by arrows. See SI Appendix: Section S3 and S4 for details on model simulation and computation of lag-1 AC. $a=1 \times 10^{-4}$, $b=-0.2$.

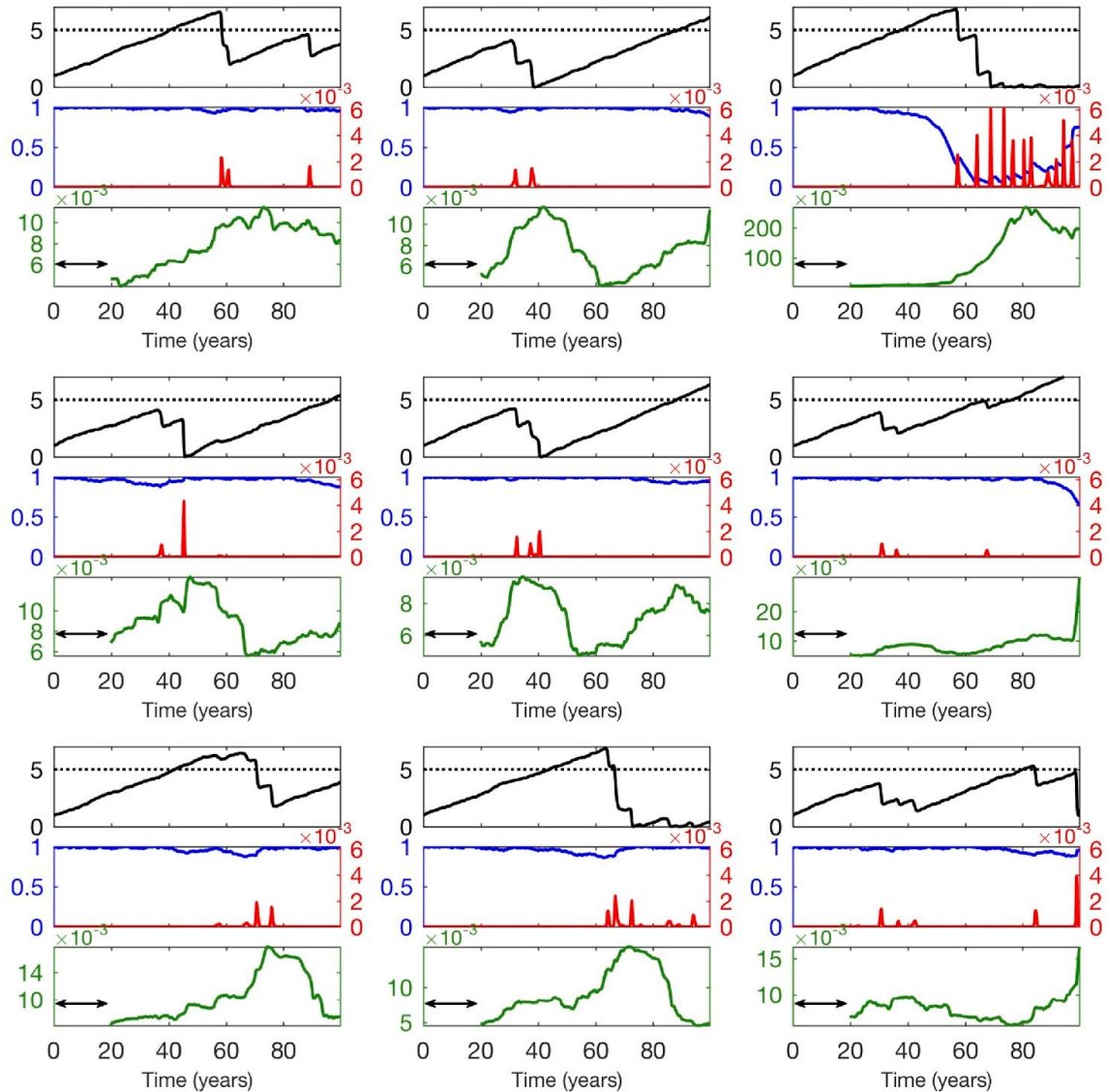


Figure S16: Adaptive dynamics in extended model where perceived risk obeys $d\omega/dt = a + bI(t)$, in nine model realizations, coefficient of variation indicator. Each set of three subpanels shows the results of a single realization. In each set, top subpanel shows time evolution of $\omega(t)$ (solid line) and location of critical transition (dashed line); middle subpanel shows vaccine uptake (blue) and infection prevalence (red); bottom subpanel shows coefficient of variation (green) of x time series computed using a rolling window of width indicated by arrows. See SI Appendix: Section S3 and S4 for details on model simulation and computation of coefficient of variation. $a=1 \times 10^{-4}$, $b=-0.2$.

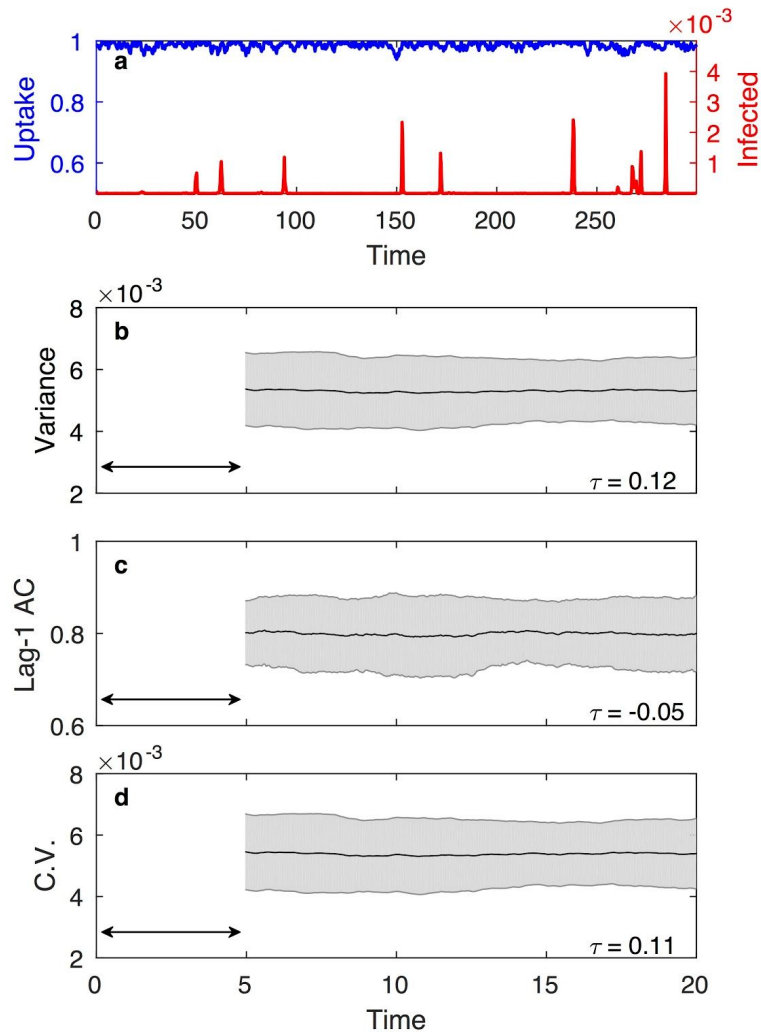


Figure S17: Model dynamics away from the critical transition do not show critical slowing down. (a) Vaccine uptake (blue) and infection prevalence (red) when perceived risk is set at $\omega=1 \times 10^{-4}$; (b) variance; (c) lag-1 autocorrelation; and (d) coefficient of variation of time periods 40 years leading up to 50 different outbreaks in model simulations. Mean is denoted by black line and shaded regions represent one standard deviation over the 50 samples. Double-headed arrows denotes width of rolling window used to compute lag-1 AC and variance. Other parameter values and details of model simulation and indicator computation appear in SI Appendix: Sections S3 and S4.