



# Strong Inference in Mathematical Modeling: A Method for Robust Science in the Twenty-First Century

Vitaly V. Ganusov<sup>1,2,3\*</sup>

<sup>1</sup> Department of Microbiology, University of Tennessee, Knoxville, TN, USA, <sup>2</sup> Department of Mathematics, University of Tennessee, Knoxville, TN, USA, <sup>3</sup> National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN, USA

While there are many opinions on what mathematical modeling in biology is, in essence, modeling is a mathematical tool, like a microscope, which allows consequences to logically follow from a set of assumptions. Only when this tool is applied appropriately, as microscope is used to look at small items, it may allow to understand importance of specific mechanisms/assumptions in biological processes. Mathematical modeling can be less useful or even misleading if used inappropriately, for example, when a microscope is used to study stars. According to some philosophers (Oreskes et al., 1994), the best use of mathematical models is not when a model is used to confirm a hypothesis but rather when a model shows inconsistency of the model (defined by a specific set of assumptions) and data. Following the principle of strong inference for experimental sciences proposed by Platt (1964), I suggest “strong inference in mathematical modeling” as an effective and robust way of using mathematical modeling to understand mechanisms driving dynamics of biological systems. The major steps of strong inference in mathematical modeling are (1) to develop multiple alternative models for the phenomenon in question; (2) to compare the models with available experimental data and to determine which of the models are *not* consistent with the data; (3) to determine reasons why rejected models failed to explain the data, and (4) to suggest experiments which would allow to discriminate between remaining alternative models. The use of strong inference is likely to provide better robustness of predictions of mathematical models and it should be strongly encouraged in mathematical modeling-based publications in the Twenty-First century.

## OPEN ACCESS

### Edited by:

Ruy Ribeiro,  
Los Alamos National Laboratory, USA

### Reviewed by:

Rustom Antia,  
Emory University, USA  
Andrew Yates,  
University of Glasgow, UK

### \*Correspondence:

Vitaly V. Ganusov  
vitaly.ganusov@gmail.com

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
Frontiers in Microbiology

**Received:** 22 April 2016

**Accepted:** 07 July 2016

**Published:** 22 July 2016

### Citation:

Ganusov VV (2016) Strong Inference in Mathematical Modeling: A Method for Robust Science in the Twenty-First Century. *Front. Microbiol.* 7:1131. doi: 10.3389/fmicb.2016.01131

**Keywords:** robust science, mathematical modeling, immunology, microbiology, public health, scientific method

## 1. THE CORE OF MATHEMATICAL MODELING

What is the use of mathematical modeling in biology? The answer likely depends on the background of the responder as mathematicians or physicists may have a different answer than biologists, and the answer may also depend on the researcher’s definition of a “model.” In some cases models are useful for estimation of parameters underlying biological processes when such parameters are not directly measurable. For example, by measuring the number of T lymphocytes over time and by utilizing a simple model, assuming exponential growth, we can estimate the rate of expansion of T cell populations (De Boer et al., 2001). In other cases, making the model

may help think more carefully about contribution of multiple players and their interactions in the observed phenomenon. In general, however, mathematical models are most useful when they provide important insights into underlying biological mechanisms. In this opinion article, I would to provide my personal thoughts on the current state and future of mathematical modeling in biology with the focus on the dynamics of infectious diseases. As a disclosure I must admit that I am taking an extreme, provocative view, based on personal experience as a reader and a reviewer. I hope that this work will generate the much needed discussion on uses and misuses of mathematical models in biology and perhaps will result in quantitative data on this topic.

In my experience, in the area of dynamical systems/models of the within-host and between-host dynamics of infectious diseases, the two most commonly given answers to the question of the “use of mathematical models” are (1) models help us understand biology better; and (2) models help us predict the impact of interventions (e.g., gene knockouts/knockins, cell depletions, vaccines, treatments) on the population dynamics. Although there is some truth to these answers the way mathematical modeling in biology is generally taught and applied rarely allows one to better understand biology. In some cases mathematical models generate predictions which are difficult or impossible to test, the latter making such models unscientific per the definition of a scientific theory according to one of the major philosophers of science in the Twentieth Century Karl Popper (Popper, 2002). Moreover, mathematical modeling may result in questionable recommendations for public health-related policies. My main thesis is that while, in my experience, much of current research in mathematical biology is aimed at finding the right model for a given biological system, we should pay more attention to understanding which biologically reasonable models do not work, i.e., are not able to describe the biological phenomenon in question. According to Karl Popper, proving a given hypothesis to be correct is impossible while rejecting hypotheses is feasible (Oreskes et al., 1994; Popper, 2002).

What is a mathematical model? In essence, mathematical model is a hypothesis regarding a phenomenon in question. While any specific model always has an underlying hypothesis (or in some cases, a set of hypotheses), the converse is not true as multiple mathematical models could be formulated for a given hypothesis. In this essay I will use words “hypothesis” and “model” interchangeably. The core of a mathematical model is the set of model assumptions. These assumptions could be based on some experimental observations or simply be a logical thought based on everyday experience. For example, for an ordinary differential equation (ODE)-based model, the assumptions are the formulated equations which include functional terms of interactions between species in the model, parameters associated with these functions, and initial conditions of the model. The utility of mathematics lies in our ability to logically follow from the assumptions to conclusions on the system’s dynamics. Thus, mathematical modeling is a logical path from a set of assumptions to conclusions. Such a logical path from axioms to theorems was termed by some as a mathematical revolution in the Twentieth Century (Quinn, 2012). However, while in mathematics it is

vital to formulate a complete set of axioms/assumptions to establish verifiable, true statements such as theorems (Quinn, 2012), a complete set of assumptions is impossible in any biology-based mathematical model due to the openness of biological systems (or any other natural system, Oreskes et al., 1994). Therefore, biological conclusions stemming from analysis of mathematical models are inherently incomplete and are in general strongly dependent on the assumptions of the model (De Boer, 2012). While such dependency of model conclusions on model assumptions may be viewed as a weakness but it is instead the most significant strength of mathematical modeling! By varying model assumptions one can vary model predictions and subsequently by comparing predictions to experimental observations, sets of assumptions which generate predictions consistent and inconsistent with the data can be identified. This is the **core of mathematical modeling** which can provide profound insights into biological processes. While it is often possible to provide mechanistic explanations for some biological phenomena from intuition—and many biologists do it—it is often hard to identify sets of implicit assumptions made during such a verbal process. Mathematical modeling by requiring one to define the model specifies such assumptions explicitly. Inherent to this interpretation of mathematical modeling is the need to consider multiple sets of assumptions (or models) to determine which are consistent and, more importantly, which are not consistent with experimental observations. Rather than a thorough expedition to test multiple alternative models, in my experience as a reader and a reviewer many studies utilizing mathematical modeling in biology have been a quest to find (and analyze) a single “correct” model.

I would argue that studies in which a single model was considered and in which the developed model was not rigorously tested against experimental data, do not provide robust biological insights (see below). Pure mathematical analysis of the model and its behavior (e.g., often performed steady state stability analyses for ODE-based models) often provides little insight into the mechanisms driving dynamics in specific biological systems. Failure to consider alternative models often results in biased interpretation of biological observations. Let me give two examples.

Discussion of predator-prey interactions in ecology often starts with the Lotka-Volterra model which is built on very simple and yet powerful basic assumptions (Mooney and Swift, 1999; Kot, 2001). The dynamics of the model can be understood analytically and predictions on the dynamics of predator and prey abundances can be easily generated. The observation of the hares and lynx dynamics in Canada has been often presented as evidence that predator-prey interactions driven the dynamics of this biological system (Mooney and Swift, 1999). While it is possible that the dynamics was driven by predator-prey interactions, recent studies also suggest that the dynamics could be driven by self-regulating factors and weather activities influencing independently each of the species (Brauer and Castillo-Chávez, 2001; Zhang et al., 2007). A more robust modeling approach would be to start with observations of lynx and hare dynamics and ask about biological mechanisms which could be driving such dynamics including predator-prey

interactions, seasonality, or both (Hilborn and Mangel, 1997). The data can then be used to test which of these sets of assumptions is more consistent with experimental data using standard model selection tools (Burnham and Anderson, 2002).

In immunology, viral infections often lead to generation of a large population of virus-specific effector CD8 T cells, and following clearance of the infection, there is formation of memory CD8 T cells (Ahmed and Gray, 1996; Kaech and Cui, 2012). However, how memory CD8 T cells are formed during the infection has been a subject of a debate (Ahmed and Gray, 1996). One of the earlier models assumed that memory precursors proliferate during the infection and produce terminally differentiated, nondividing effector T cells, which then die following clearance of the infection (Wodarz et al., 2000; Bocharov et al., 2001; Wodarz and Nowak, 2002; Fearon et al., 2006). While this model was used to explain several biological phenomena, later studies have shown that this model failed to accurately explain experimental data on the dynamics of CD8 T cell response to lymphocytic choriomeningitis virus (Antia et al., 2005; Ganusov, 2007). More precisely, the model was able to accurately fit experimental data but it required unphysiologically rapid interdivision time for activated CD8 T cells [e.g., 25 min in Ganusov (2007)] which was inconsistent with other measurements made to date. Constraining the interdivision time to a larger value (e.g., 3 h) resulted in a poor model fit of the data. Therefore, development of adequate mathematical models cannot be all based on “basic principles” and must include comparison with quantitative experimental data.

These examples illustrate how mathematical modeling can teach us about mechanisms underlying biological processes. When a model is developed using some basic biological assumptions/mechanisms and yet such a model is unable to accurately describe quantitative biological data, we learn something. We learn that the mechanisms that we thought should be important in explaining the phenomenon are incorrect (or that we modeled them incorrectly). In this case, modeling provides important information that some aspects of biology that we thought we knew we actually do not know. In the case of memory CD8 T cell differentiation, the poor assumption was that effector T cells do not proliferate (Ganusov, 2007). An alternative situation is when it is believed that only one mechanism explains a biological phenomenon, and yet several different models can be formulated and all models are able to accurately describe experimental data. Again, such a result would illustrate that specific data can be explained by more than one mechanism and additional experiments are needed to further discriminate between alternative models. Although this has not been formally done, two alternative mechanisms (predator-prey and seasonality) may be reasonable explanations of the hare-lynx dynamics in Canada.

## 2. STRONG INFERENCE IN MATHEMATICAL MODELING

Strong inference was proposed over 50 years ago to promote rapid science (Platt, 1964). Platt suggested that despite

a commonly spread “...polite fiction that all science is equal... some areas of science progress faster than others” (Platt, 1964). Platt (1964) proposed that by choosing well formulated questions and hypotheses and by designing discriminatory experiments, one can progress faster with understanding of the underlying phenomena. According to strong inference, the following steps must be taken to investigate a given scientific question (Platt, 1964):

1. Devising alternative hypotheses;
2. Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;
3. Carrying out the experiment so as to get a clean result;
- 1'. Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain; and so on.

These recommendations were highly influential as judged by the number of citation (1439 in Web of Science or 2867 in Google scholar as of April 5th, 2016); however, it does not appear that they have been widely adopted in biological sciences (Jewett, 2005). Two major points of these recommendations include (1) formulation of a set of alternative hypotheses and (2) attempt to reject, not to confirm, these hypotheses. The idea of formulating multiple hypotheses goes back to another important paper on “The method of multiple working hypotheses” (Chamberlin, 1890) which recently received an update (Elliott and Brook, 2007). The idea of testing hypotheses to reject them goes back to Karl Popper, who proposed that falsification of hypotheses is the core of the scientific method (Popper, 2002). Strong inference received its share of criticism suggesting that it cannot be applied in some areas of research and that it does not promote rapid science (O’Donohue and Buchanan, 2001). Indeed, testing  $n > 1$  multiple hypotheses is unlikely to provide rapid progress because it would probably take  $n$  times longer to find the answer as compared to that if there were only one hypothesis to start with. However, strong inference will likely result in more robust results than results based on a single hypothesis, and therefore, overall, multiple hypotheses-driven research provides more rapid progress for the field as it cuts out early wrong leads. One author suggested that the use of strong inference may occur more frequently in industry than in academia due to a higher focus of industrial research on robustness rather than novelty (Ehlers, 2016). Robust conclusions rather than novel results are also viewed as a feature of good scientists both by general public and professional researchers (Ebersole et al., 2016).

In my view, not all mathematical modeling studies are equal and some provide better insights into biological mechanisms than others. By extending Platt’s ideas to mathematical modeling I propose the following steps for “strong inference in mathematical modeling” in biology:

1. For a given biological question and associated experimental data, formulate several alternative mathematical models aimed at explaining the data;
2. Compare model predictions with experimental data with the goal of excluding as many of the alternative models as possible;

3. For the rejected models, determine reasons why the models were not able to accurately describe the data;
4. For the models that are consistent with the data, generate predictions for experiments which would allow one to discriminate between these alternative models;
- 1'. As new data are available, recycle the procedure by making sub-models, alternative models, and so on.

To avoid misinterpretation two issues must be explained further: what different models are and what it means to reject a model.

There are two levels at which alternative models can be defined. One is the basic/core mechanism of the mathematical model and another is specific model formulations within such a core mechanism. Using hare-lynx dynamics as an example, two core mechanisms could include predator-prey interactions or season-driven dynamics. (Perhaps the reader already came up with a third core mechanism?) Using a given specific core mechanism one now can write different formulations of the model, for example, how predator consumes the prey and how the prey biomass translates into predator biomass. Multiple formulations are possible and these all are alternative models, and yet they all have the same basic core mechanism. In essence, the model core is an equivalent of the main hypothesis responsible for the observed phenomenon. Similarly, seasonality can enter the model directly assuming time-dependent birth/death rates of hares and lynx or indirectly by assuming time-dependent variability in resources. These formulations also can be viewed as alternative models. Rejection of a specific mathematical model does not necessarily invalidate the core mechanism but rejection of a set of alternative models based on a given core mechanism will raise doubts whether such a core mechanism is responsible for the observed phenomenon. The best use of strong inference is a rejection of a core mechanism.

Criteria of model rejection are not well established and rejection can be done on absolute or relative grounds. When comparing model predictions and data one could ask if the model is adequately describing the data. Two tests could be of particular importance such as goodness of fit test and lack of fit test (Bates and Watts, 1988). These tests require data with sufficient richness but in some cases, incompatibility between model and data can be determined (Noecker et al., 2015). When using a set of alternative models other tests such as likelihood ratio test or information criteria (AIC, BIC, etc.) can be also used (Bates and Watts, 1988; Burnham and Anderson, 2002; Johnson and Omland, 2004) to determine which of the models are less likely to be consistent with the data. Similarly, comparison with data may allow to reject a core mechanism or more commonly, reject specific formulations of the core mechanism. Issues associated with identifiability of mathematical models and precise estimation of model parameters in some case may not allow to reject specific models (Meshkat et al., 2009; Raue et al., 2009).

Proper application of strong inference in mathematical modeling depends critically on choosing a “good” question which has only a limited number of possible core mechanisms. It is clear that “big” fundamental questions often have many potential answers (O’Donohue and Buchanan, 2001) and from the perspective of strong inference, big questions can rarely be

exhaustively explored. As continuous application of the method of multiple working hypotheses “develops a habit of parallel or complex thought” (Chamberlin, 1890), continuous application of strong inference allows development of a skill of asking the “good” questions and recognition when asked questions are “bad.”

As the method of multiple working hypotheses has a “danger of vacillation” (Chamberlin, 1890), strong inference may fail when none of the alternative models can be rejected. In fact, it has been argued that inability to reject hypotheses/models may be a feature of ecological studies (Hobbs and Hilborn, 2006). One proposed solution is to use model averaging where predictions of different models are “weighted” based on the models’ consistency with experimental data (Hoeting et al., 1999; Burnham and Anderson, 2002). Model averaging is not without problems, however, including situations where alternative models generate contradictory predictions (Grueber et al., 2011). In my view, inability to apply principles of strong inference to reject some of the alternative models indicates two potential problems: (1) the data are poor and insufficient to discriminate between alternative models (so more and better data need to be collected), and (2) the formulated question is “bad” (so a better formulated question is needed).

One useful example of the use of strong inference comes from the analysis of movement patterns of activated CD8 T cells in murine brains (Harris et al., 2012). Using intravital imaging the authors recorded coordinates of T cells in the brain over long periods of time. By comparing predictions of multiple mathematical models the authors concluded that only one in the list of several alternative models, based on generalized Levy walks, could explain all data with reasonable quality (Harris et al., 2012). Future studies utilizing further strong inference would need to discriminate between cell-intrinsic vs. environment-driven core mechanisms explaining this type of walk of T cells in the brain.

With principles of strong inference the power of mathematical modeling can be truly revealing. Closer collaborations between experimentalists and modelers leading to discrimination between alternative models using data would likely result in substantial robust gains in our understanding of biological processes.

### 3. DANGERS OF SINGLE HYPOTHESIS/MODEL-DRIVEN RESEARCH

While scientific benefits of multiple hypotheses/models-driven research are hard to deny, dangers of using single hypotheses in research have not been widely emphasized. Already in 1890, Chamberlin (1890) warned about biases resulting from “dominant theory” or “single hypothesis”-driven research and why thinking in terms of multiple hypotheses must extend beyond science and be common practice for everyone in the world. I would like to present three examples, in which single hypothesis/mathematical model-driven research limits and sometimes biases our understanding of biology. These examples represent my hypothesis on limited robustness of single

mathematical model-based studies; this hypothesis will have to be tested and perhaps rejected in the future.

### 3.1. Biased Predictions

One of the virtues of mathematical models is often cited their predictive power. Indeed, mathematical models are used to make predictions in many areas of science including biology. The types of models used to make predictions vary in their complexity from simple, few equations-based models to models including hundreds of variables. How robust are predictions of such models? My thesis is that predictions based on a single mathematical model are unlikely to be robust (De Boer, 2012).

Recently, Evans et al. (2013) questioned whether general, very simple models are useful in making quantitative predictions on vital, public-health related issues. The authors argued that such general models by design are relatively simple and are aimed at describing as many situations as possible. The authors also argued that models that are designed for specific systems and parameterized from specific experimental data, are likely to be more precise in predictions. Such case-specific models are thought to be more useful in guiding policies for control of infectious diseases (Evans et al., 2013). The authors illustrated their point by discussing the predictions of two mathematical models on the level of vaccination required to eradicate rabies in the fox populations in Europe (Anderson et al., 1981; Eisinger and Thulke, 2008). Evans et al. (2013) argued that simple, susceptible-infected-recovered mathematical model overestimated the level of vaccination needed for rabies eradication (Anderson et al., 1981). Such a simple model predicted that 70% of foxes had to be vaccinated for efficient control. A more complex model, including details of the local spread of the infection from rabid to susceptible foxes, predicted a lower vaccination level of 60% (Eisinger and Thulke, 2008). Although such a 10% difference may appear small, Eisinger and Thulke (2008) suggested that the vaccination campaign based on the prediction of the simple model may have cost over several millions of euros more than was needed. The authors concluded that in order to make public health-related predictions for a specific biological system, the models should include sufficient detail about that system so the model predictions are accurate and precise (Evans et al., 2013). Thus, predictions of a single model may not be robust, and in some cases, predicted interventions may cost more than needed.

Another example comes from early predictions of potential size of the Ebola virus epidemics in Africa in 2014–2016 (Butler, 2014). Initial studies by considering simple models predicted devastating impact of the epidemic on human population which luckily did not occur (Butler, 2014; Pandey et al., 2014). Later analyses revealed that simple models were inadequate by ignoring potential heterogeneity in behavior which translated into large variability in transmission efficacy (Drake et al., 2015). Although there is a consensus that mathematical modeling is needed to understand biological phenomena including epidemiology of infectious diseases (Lofgren et al., 2014), non-robust model predictions which overestimate risks are perhaps even more harmful than models that underestimate the risks. In fact, good modeling practice is in general to provide minimal estimates of

the risk. Examples of wrong predictions may fuel unwarranted public debate on trustworthiness of mathematical models, for example, predicting climate change. Taken together, studies that are based on the analysis of a single model are not expected to produce robust predictions (Oreskes et al., 1994). Predictive studies illustrating which alternative models have been considered in the analysis, which models have been rejected and why, and whether predictions of the remaining models are self-consistent, will lead to robust predictions and should be encouraged.

### 3.2. Unreproducible Science

The great feature of science is its self-correcting nature. Some theories have persisted for decades but have been shown later to be incorrect as new ideas and data accumulated. While exceptions clearly exist and there are still common myths despite experimental evidence otherwise (Scudellari, 2015), science has been mostly self-correcting. I would argue that in some cases consideration of a single hypothesis and failure to consider and reject alternatives has caused dominance of an eventually wrong theory. In some cases, self-correction in sciences took long time with resources wasted and lives affected. One example is on the development of understanding of motions of planets with a complete dominance of Ptolemy's theory of immotile Earth with Sun and planets moving in circular orbits (Danielson and Graney, 2014). If Tycho Brahe, one of the major astronomers collecting data to support Ptolemy's circular orbits-based theory, and other scientists at the time considered alternatives of elliptic circles and movable Earth, perhaps science would progress faster, reach more robust conclusions, and Bruno and Galileo would not have suffered (Danielson and Graney, 2014). There is more recent, perhaps an extreme example of a crime conviction of an innocent person based on consideration of a single hypothesis (Nuzzo, 2015).

The common practice of considering a single hypothesis and collecting data to "prove" it can bias interpretation and may result in unreproducible results. In recent years it has been noted by several groups of investigators that many of the results in biological sciences are unreproducible (Prinz et al., 2011; Begley and Ellis, 2012; Collaboration, 2015; Freedman and Gibson, 2015; Freedman et al., 2015). In particular, biotech company Amgen attempted to reproduce 53 "landmark" papers from cancer biology and was able to reproduce only 6 (Begley and Ellis, 2012). Overall, a recent review suggests that at least 50% of reanalyzed studies are unreproducible (Freedman et al., 2015). If these findings can be extrapolated to the whole field of biomedical research one study estimates that over \$28B are wasted on unreproducible studies, and half of those expenditures are suggested to result from inappropriate study design and data analysis (Freedman et al., 2015).

It remains unknown whether reproducibility of mathematical modeling-based studies is different from that of science in general (or biology in particular, Boulesteix et al., 2015). For example, one recent study could reproduce less than half of bioinformatic analyses of published microarray gene expression data (Ioannidis et al., 2009). The definition of reproducibility may be difficult in general as it may vary by researcher (Goodman et al., 2016). For

one type of mathematical modeling studies which do not involve any experimental data we generally expect full reproducibility if the authors correctly wrote and analyzed their model and/or appropriately simulated its dynamics. However, programming errors may still occur. A lower level of reproducibility may be expected for studies utilizing both mathematical models and analysis of experimental data. I analyzed a subset of data from a recent survey by Nature (Baker, 2016) by focusing on responses by scientists from the field of “Biology” with expertise in “Bioinformatics and Computational Biology” ( $n_1 = 36$ ) or “Systems Biology” ( $n_2 = 9$ ,  $n = n_1 + n_2 = 45$  surveys in total). I found that computational biologists are at least as skeptical about the state of reproducibility of studies in their fields as compared to all scientists surveyed. In particular, computational biologists believe that on average only 50% of studies in their field are reproducible (compared to 58% for general population, Mann-Whitney test,  $p = 0.02$ ), 27% believe that computational biology has similar level of reproducibility compared to other fields (vs. 21% for all scientists,  $\chi^2_{(1)} = 0.76$ ,  $p = 0.38$ ), and 73% of computational biologists believe that failure to reproduce results is the major problem in the field (as compared to 59% of all scientists surveyed,  $\chi^2_{(1)} = 3.85$ ,  $p = 0.05$ ). Interestingly, 20% of computational biologists were told that someone could not reproduce their work (vs. 18% for all scientists,  $\chi^2_{(1)} = 0.12$ ,  $p = 0.73$ ). Thus, there is a general concern about the level of reproducibility of mathematical modeling-based studies.

A large number of unreproducible studies is paralleled by a recent increase in percentage of retracted peer-reviewed papers (Fang et al., 2012; Grieneisen and Zhang, 2012; Fanelli, 2013; Castillo, 2014). While increased scrutiny of published papers may have contributed to the rise in the number of retracted articles (Fanelli, 2013), the increased competition in research, especially in biomedical sciences, leading to the “publish-or-perish” culture is a very like cause for the growing number of unreproducible studies and retracted papers (Steen et al., 2013). The number of retracted mathematical modeling-based papers remains relatively low (a simple search for “mathematical model” on RetractionWatch.com yielded under ten hits as of April 5th, 2016).

The need for more robust ways of doing science, including mathematical modeling, is well recognized (Begley and Ellis, 2012; Fang and Casadevall, 2012). By focusing mathematical modeling analyses on a single model and by showing qualitative consistency of the model and data we commit a cognitive/confirmation bias (Kaptchuk, 2003; Editorial, 2015). Confirmation bias appears to be widespread in the mathematical modeling literature where consistency of a model with experimental observations occurs much more frequently than rejection of models. Even in cases when model predictions match qualitatively other, potentially independent data, there is a risk of so-called “therapeutic illusion” (Casarett, 2016), an inability to recognize that alternative mechanisms, not included in the model, could explain additional data too. Several suggestions have been made to improve reproducibility and robustness of science including use of strong inference (Nuzzo, 2015), improved trainings (Moher and Altman,

2015), performing blind analyses of the data (MacCoun and Perlmutter, 2015), the need for independent analyses of the same data/models by different teams prior to publication (Silberzahn and Uhlmann, 2015), and standardization of tools (Baker, 2015). There is also a need to reduce overoptimistic reporting in mathematical modeling-based studies (Boulesteix, 2015) and reduce uncertainties in predictions of mathematical models (Kirk et al., 2015). The use of principles of strong inference should increase robustness of predictions of mathematical models and in general, should reduce the amount of unreproducible research in biology.

### 3.3. Development of Large Models

The formulation and analysis of multiple alternative mathematical models can clearly increase robustness of conclusions and improve our ability to make accurate predictions. Robustness of predictions of mathematical models for public health-related policies is particularly important. To avoid the need to formulate multiple alternative models for a given phenomenon researchers often construct models that include many of known mechanisms in the biological system of interest. Such a model is then expected to be able to explain a large number of different phenomena, and there is a hope that at some choice of parameters the model behavior will capture true biological forces at play. Such a model is viewed as useful to make specific predictions of the impact of interventions on population dynamics (Bru and Cardona, 2010; Cilfone et al., 2015). This trend for “systems” view on biological phenomena is becoming more popular and it is now being questioned whether simple models which include only a few major details about biological system are useful in making relevant forecasts (Evans et al., 2013). One of the major problems of large and complex models is that by including many mechanisms and details these models become as complex as phenomena they are trying to explain precluding detailed understanding of such models. Furthermore, by including multiple details such large models can rarely if ever be rejected which essentially makes them unscientific per Karl Popper (Popper, 2002; Ellis and Silk, 2014).

Large complex models are often compared to data to illustrate their plausibility. However, with tens to hundreds of parameters complex models can easily explain one or several datasets. Such model overfitting of the data should never be viewed as model confirmation (Oreskes et al., 1994). Only few parameters are needed to generate complex patterns as famous saying states: “with four parameters I can fit an elephant, and with five, I can make him wiggle his trunk” (Mayer et al., 2010; Ditlev et al., 2013). Development of large, complex models can be useful if such models show inconsistency of specific mechanisms with sets of experimental observations. Predictions of large models should be treated with caution unless it has been established which alternative models/mechanisms have been rejected during model development (Oreskes et al., 1994). Iterative process of model development, testing, and calibration using sufficiently extensive datasets may result in large mathematical models of robust predictive power; mathematical models predicting weather are one good example (Bauer et al., 2015). Yet, even well calibrated

weather prediction models have reasonable accuracy only for relatively short-term predictions (Bauer et al., 2015).

#### 4. CHANGING TRAINING IN MATHEMATICAL BIOLOGY

Given intuitive benefits of multiple models-driven research it is perhaps strange to realize that it remains quite rare. In part this is due to widely adopted approach to find models which explain phenomena. I believe that “the approach to find the right model” starts very early in education of a mathematical biologist, probably during undergraduate or early graduate career. Many of the classical textbooks on mathematical modeling in biology have a similar theme: (1) identify a biological problem, (2) develop a mathematical model for the problem; the degree of complexity of the model should depend on the complexity of the problem and/or underlying biology, (3) analyze the model; (4) draw the conclusions from the model behaviors and extrapolate the conclusions to the actual biological system (Segel, 1984; Mooney and Swift, 1999; Kot, 2001; Ellner and Guckenheimer, 2006; Vries et al., 2006; Percus, 2012). In this approach the developed model is often treated as a very good representation of the actual biological system and rarely the basic assumptions of the model are challenged. Education in physics and engineering proceeds in a similar fashion where complex mathematical models are derived from basic principles which are accepted to be true either because of some fundamental experiments or simply because of intuition. This approach, although being relatively straightforward, fosters an impression that if one starts with a good set of assumptions this will lead to a model which should not be questioned. Experimental data are often brought as support of the model, and when the model predictions are consistent with some, often qualitative data, the model appears to be a strong reflection of the reality (Simberloff, 2007). However, rarely the basic feature of mathematical models—that predictions are the direct consequences of the model assumptions—is investigated thoroughly by identifying model assumptions which are most critical for the “consistency” between the model and experimental observations, and which assumptions would allow the model to “fail” at explaining the data. Furthermore, in many cases consistency between models and data is indicated by qualitative or semi-quantitative comparison which does not allow to investigate in a rigorous sense whether the model is indeed an accurate enough representation of the data (Jin et al., 1999; Wang et al., 2015).

While many methods are likely to improve robustness of mathematical modeling-based (and other scientific) studies, the widespread use of strong inference is likely to be important in this endeavor (Nuzzo, 2015). Design of multiple alternative models forces the researcher to deeply understand the underlying biological question and not be satisfied with standard answers that “this is well known” but to require solid experimental support for major model assumptions. *Education of future generations of students in mathematical modeling should focus more on deeper understanding of biological details and on investigating which aspects of their models could be wrong.* If we

substitute “theory” with “model,” it was very nicely said by Ellis and Silk (2014) that research often “boils down to clarifying one question: what potential observational or experimental evidence is there that would persuade you that the theory is wrong and lead you to abandoning it? If there is none, it is not a scientific theory.” Finding boundaries when the model “breaks” at explaining the phenomenon in question would reveal limitations of the model and of its predictions. *Therefore, future mathematical modelers should be able to understand details of biological experiments, how the data are collected and analyzed, so such data are used with most efficiency for model development and testing. Such training thus must extend beyond traditional education in mathematics, engineering, and computer science.*

One of the major difficulties with multiple models-driven research and strong inference is to identify the number of alternative models/hypotheses one needs to consider to satisfy principles of strong inference (Platt, 1964). Choosing a “good” question is key in this process. Wise application of strong inference requires selection of “good” questions for which only a limited number of alternative hypotheses (or core mechanisms) exist (Platt, 1964). Choosing the “good” question is an endeavor and skill on its own; it is a part of scientific method and it requires specific training. *Education in mathematical modeling should focus more on developing skills on identifying biological problems which have a limited number of possible answers and which can be addressed using mathematical modeling.* For example, if one finds too many alternative explanations for his/her question, perhaps he/she is not asking a “good” question. In practice, consideration of two or more models would be likely to be better than study with a single model, and formulation and analysis of models with alternative core mechanisms is most preferable per strong inference.

It has to be realized that predictions of any single model for a biological system are not likely to be robust due to inherent openness of biological systems (Oreskes et al., 1994). Therefore, any single model is very limited in its use. However, a collection of alternative models is more likely to generate robust predictions; alternatively, analysis of such models could suggest inability to make robust predictions due to lack of appropriate data to reject alternative models. In this case, such multiple models-driven analysis may suggest areas for further experimental investigations. The idea of limited robustness of mathematical models in describing biological phenomena needs to be percolated in educational curriculum of undergraduate and graduate students, and this notion needs to be more widely stated in the professional modeling community. Realization that for every biological problem there are likely several alternative mechanisms/models needs to be eventually translated in research where it is not acceptable anymore to have a publication with only one mathematical model analyzed. We need to see mathematical biology research to move to the stage where in most publications the authors propose multiple models and discriminate between these models using quantitative biological data. *Education of future generation of mathematical modelers must include training in building of alternative mathematical models and in techniques to discriminate between alternative models using experimental data* (Burnham and Anderson, 2002; Johnson and Omland,

2004). When presented with results from a mathematical modeling-based study we should always ask **the question** (adapted from Platt, 1964): “But Sir/Madam, which mathematical models/mechanisms have you rejected in your study?”

*Training of a new generation of scientists in mathematical biology should involve more reading and discussion of the basics of scientific method.* Three papers are of particular importance and they should form the core of the graduate curriculum in graduate schools and specifically, of programs on mathematical modeling (Chamberlin, 1890; Platt, 1964; Oreskes et al., 1994). While I have discussed the ideas of the papers by Chamberlin (1890) and Platt (1964), an essay by Oreskes et al. (1994) clearly defined usefulness and limitations of mathematical modeling of open natural systems. In particular, the authors strongly cautioned against use of the words “verification” and “validation” to indicate “quality” of mathematical models as these terms exaggerate the limited ability of models to make robust predictions. In fact, “verification” of models is impossible per word definition due to the openness of natural systems, and in most cases the use of the word “validation” is synonymous to “verification” and thus is also inappropriate. The authors discussed in detail why verification/validations of models (or any logical statement) is impossible in natural sciences, and highlighted many philosophical developments on the nature of scientific method in the early Twentieth Century that is rarely discussed in graduate programs nowadays.

An important component of learning about mathematical modeling in biology is a realization that good modeling requires good understanding of the developed mathematical models. When does one understand the model, in a true sense of understanding? I believe that for simple models with a few parameters, true understanding is realized when one intuitively can predict the impact of the change in a model parameter or a combination of parameters on the model dynamics. Such a detailed understanding of the model also allows for insights in situations when the model is not able to fit/describe experimental data—i.e., why isn't the model able to explain experimental data? What is wrong with it? Deeper understanding of the model can point to parts of the model that are responsible for such discrepancy. Intuitive understanding of the model is very difficult or impossible for models with tens to hundreds of parameters. Yet, such an understanding is needed if the model fails to explain well some experimental data. How can one understand such a model? The traditional approach for understanding complex models is sensitivity analysis (Marino and Kirschner, 2004). Sensitivity analysis can allow to rank parameters of the model or the combination of parameters in terms of their impact on behavior of specific model components, e.g., density of species at some time point. I would argue, however, that in many cases sensitivity analyses do not give a good understanding of the model behavior because answers may depend on the method used and because sensitivity analysis often does not specify why this and not another parameter is the most important in the model dynamics. However, analyses which provide rational explanations of why specific parameters or parameter combinations drive model dynamics will likely reveal relative importance of different biological mechanisms.

*Education of future mathematical modelers should include basics of sensitivity analyses and understanding when such analyses are informative and when they are not.*

## 5. CONCLUSIONS

A simple and effective critique of multiple hypotheses/models-driven research is to make counter examples of studies utilizing a single mathematical model and yet providing important biological insights. For instance, very well known studies utilizing a single ODE-based mathematical model estimated the rate of turnover of HIV and HIV-infected cells (Ho et al., 1995; Wei et al., 1995). Although the success of this pioneering work to accurately estimate the life-span of infected cells is well known, the failure of the model to accurately predict turnover of CD4 T cells due to incorrect assumption of CD4 T cell recovery due to production of new T cells is rarely acknowledged (Ho et al., 1995; Pabst and Rosenberg, 1998; Bucy et al., 1999). Furthermore, because we tend to remember “winners” and forget “losers,” it is very likely that many predictions of single mathematical modeling-based studies are incorrect or not robust to changes in the model assumption. It would be useful to generate data on the frequency of “correct” vs. “incorrect” predictions of studies based on single vs. multiple mathematical models although it may be difficult to define “correctness” of predictions.

Even in the absence of such data I propose that in order for mathematical modeling to become more robust, more practical and relevant for infectious disease biology we, mathematical modelers, need to re-think how we do research and how we train new generations of students. It is possible that the current format in which students, taking mathematical modeling in biology courses, get exposed to sets of standard models and their properties needs to be changed to observation-driven training where students develop models to explain particular experimental observations. Basic biological principles can be used to drive the development of models with variable levels of complexity and models the alternative mechanisms. Comparison to quantitative experimental data then can be used to test which of the models (i.e., mechanisms) are not consistent with the data and why (Popper, 2002).

Given that mathematical models are increasingly playing an important role in policy decision making (Christley et al., 2013), it is the time to change the way many mathematicians approach modeling, and we need to change the way we teach mathematical modeling at universities. Devising as many as possible alternative models for every biological question and comparing model predictions with quantitative experimental data to reject the models will allow mathematical modeling to become a scientific procedure generating more robust predictions.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## ACKNOWLEDGMENTS

I would like to thank many colleagues who have influenced my view on mathematical modeling including Rustom Antia, Rob De Boer, Alan Perelson, past and

present members of my group, Sivan Leviyang for comments on earlier versions of the paper, and reviewers who raised important points on uses of mathematical modeling. This work was in part supported by the AHA grant to VG.

## REFERENCES

- Ahmed, R., and Gray, D. (1996). Immunological memory and protective immunity: understanding their relation. *Science* 272, 54–60. doi: 10.1126/science.272.5258.54
- Anderson, R. M., Jackson, H. C., May, R. M., and Smith, A. M. (1981). Population dynamics of fox rabies in Europe. *Nature* 289, 765–771. doi: 10.1038/289765a0
- Antia, R., Ganusov, V. V., and Ahmed, R. (2005). The role of models in understanding CD8+ T-cell memory. *Nat. Rev. Immunol.* 5, 101–111. doi: 10.1038/nri1550
- Baker, M. (2015). Reproducibility crisis: blame it on the antibodies. *Nature* 521, 274–276. doi: 10.1038/521274a
- Baker, M. (2016). Is there a reproducibility crisis? *Nature* 533, 452–454. doi: 10.1038/533452a
- Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1002/9780470316757
- Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature* 525, 47–55. doi: 10.1038/nature14956
- Begley, C. G., and Ellis, L. M. (2012). Drug development: raise standards for preclinical cancer research. *Nature* 483, 531–533. doi: 10.1038/483531a
- Bocharov, G., Klenerman, P., and Ehl, S. (2001). Predicting the dynamics of antiviral cytotoxic T-cell memory in response to different stimuli: cell population structure and protective function. *Immunol. Cell. Biol.* 79, 74–86. doi: 10.1046/j.1440-1711.2001.00985.x
- Boulesteix, A.-L. (2015). Ten simple rules for reducing overoptimistic reporting in methodological computational research. *PLoS Comput. Biol.* 11:e1004191. doi: 10.1371/journal.pcbi.1004191
- Boulesteix, A.-L., Stierle, V., and Hapfelmeier, A. (2015). Publication bias in methodological computational research. *Cancer Inform.* 14(Suppl 5), 11–19. doi: 10.4137/CIN.S30747
- Brauer, F., and Castillo-Chávez, C. (2001). *Mathematical Models in Population Biology and Epidemiology, Texts in Applied Mathematics*. New York, NY: Springer.
- Bru, A., and Cardona, P.-J. (2010). Mathematical modeling of tuberculosis bacillary counts and cellular populations in the organs of infected mice. *PLoS ONE* 5:e12985. doi: 10.1371/journal.pone.0012985
- Bucy, R. P., Hockett, R. D., Derdeyn, C. A., Saag, M. S., Squires, K., Sillers, M., et al. (1999). Initial increase in blood CD4(+) lymphocytes after HIV antiretroviral therapy reflects redistribution from lymphoid tissues. *J. Clin. Invest.* 103, 1391–1398. doi: 10.1172/JCI5863
- Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, NY: Springer-Verlag.
- Butler, D. (2014). Models overestimate Ebola cases. *Nature* 515:18. doi: 10.1038/515018a
- Casarett, D. (2016). The science of choosing wisely—overcoming the therapeutic illusion. *New Engl. J. Med.* 374, 1203–1205. doi: 10.1056/NEJMp1516803
- Castillo, M. (2014). The fraud and retraction epidemic. *AJNR Am. J. Neuroradiol.* 35, 1653–1654. doi: 10.3174/ajnr.A3835
- Chamberlin, T. C. (1890). The method of multiple working hypotheses: with this method the dangers of parental affection for a favorite theory can be circumvented. *Science* 15, 92–96.
- Christley, R. M., Mort, M., Wynne, B., Wastling, J. M., Heathwaite, A. L., Pickup, R., et al. (2013). “Wrong, but useful”: negotiating uncertainty in infectious disease modelling. *PLoS ONE* 8:e76277. doi: 10.1371/journal.pone.0076277
- Cilfone, N. A., Ford, C. B., Marino, S., Mattila, J. T., Gideon, H. P., Flynn, J. L., et al. (2015). Computational modeling predicts IL-10 control of lesion sterilization by balancing early host immunity-mediated antimicrobial responses with caseation during *Mycobacterium tuberculosis* infection. *J. Immunol.* 194, 664–677. doi: 10.4049/jimmunol.1400734
- Collaboration, O. S. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- Danielson, D., and Graney, C. M. (2014). The case against Copernicus. *Sci. Am.* 310, 72–77. doi: 10.1038/scientificamerican0114-72
- De Boer, R. J. (2012). Which of our modeling predictions are robust? *PLoS Comput. Biol.* 8:e1002593. doi: 10.1371/journal.pcbi.1002593
- De Boer, R. J., Oprea, M., Antia, R., Murali-Krishna, K., Ahmed, R., and Perelson, A. S. (2001). Recruitment times, proliferation, and apoptosis rates during the CD8(+) T-cell response to lymphocytic choriomeningitis virus. *J. Virol.* 75, 10663–10669. doi: 10.1128/JVI.75.22.10663-10669.2001
- Ditlev, J. A., Mayer, B. J., and Loew, L. M. (2013). There is more than one way to model an elephant. Experiment-driven modeling of the actin cytoskeleton. *Biophys. J.* 104, 520–532. doi: 10.1016/j.bpj.2012.12.044
- Drake, J. M., Kaul, R. B., Alexander, L. W., O’Regan, S. M., Kramer, A. M., Pulliam, J. T., et al. (2015). Ebola cases and health system demand in Liberia. *PLoS Biol.* 13:e1002056. doi: 10.1371/journal.pbio.1002056
- Ebersole, C. R., Axt, J. R., and Nosek, B. A. (2016). Scientists’ reputations are based on getting it right, not being right. *PLoS Biol.* 14:e1002460. doi: 10.1371/journal.pbio.1002460
- Editorial (2015). Let’s think about cognitive bias. *Nature* 526:163. doi: 10.1038/526163a
- Ehlers, M. D. (2016). Lessons from a recovering academic. *Cell* 165, 1043–1048. doi: 10.1016/j.cell.2016.05.005
- Eisinger, D., and Thulke, H.-H. (2008). Spatial pattern formation facilitates eradication of infectious diseases. *J. Appl. Ecol.* 45, 415–423. doi: 10.1111/j.1365-2664.2007.01439.x
- Elliott, L. P., and Brook, B. W. (2007). Revisiting chamberlin: multiple working hypotheses for the 21st century. *BioScience* 57, 608–614. doi: 10.1641/B570708
- Ellis, G., and Silk, J. (2014). Defend the integrity of physics. *Nature* 516, 321–323. doi: 10.1038/516321a
- Ellner, S. P., and Guckenheimer, J. (2006). *Dynamics Models in Biology*. Princeton, NJ: Princeton University Press.
- Evans, M. R., Grimm, V., Johst, K., Knuutti, T., de Langhe, R., Lessells, C. M., et al. (2013). Do simple models lead to generality in ecology? *Trends Ecol. Evol.* 28, 578–583. doi: 10.1016/j.tree.2013.05.022
- Fanelli, D. (2013). Why growing retractions are (mostly) a good sign. *PLoS Med.* 10:e1001563. doi: 10.1371/journal.pmed.1001563
- Fang, F. C., and Casadevall, A. (2012). Reforming science: structural reforms. *Infect. Immun.* 80, 897–901. doi: 10.1128/IAI.06184-11
- Fang, F. C., Steen, R. G., and Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17028–17033. doi: 10.1073/pnas.1212247109
- Fearon, D., Carr, J., Telaranta, A., Carrasco, M., and Thaventhiran, J. (2006). The rationale for the IL-2-independent generation of the self-renewing central memory CD8+ T cells. *Immunol. Rev.* 211, 104–118. doi: 10.1111/j.0105-2896.2006.00390.x
- Freedman, L. P., Cockburn, I. M., and Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biol.* 13:e1002165. doi: 10.1371/journal.pbio.1002165
- Freedman, L. P., and Gibson, M. C. (2015). The impact of preclinical irreproducibility on drug development. *Clin. Pharmacol. Ther.* 97, 16–18. doi: 10.1002/cpt.9
- Ganusov, V. V. (2007). Discriminating between different pathways of memory CD8+ T cell differentiation. *J. Immunol.* 179, 5006–5013. doi: 10.4049/jimmunol.179.8.5006

- Goodman, S. N., Fanelli, D., and Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Sci. Transl. Med.* 8, 1–6. doi: 10.1126/scitranslmed.aaf5027
- Grieneisen, M. L., and Zhang, M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *PLoS ONE* 7:e44118. doi: 10.1371/journal.pone.0044118
- Grueber, C. E., Nakagawa, S., Laws, R. J., and Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *J. Evol. Biol.* 24, 699–711. doi: 10.1111/j.1420-9101.2010.02210.x
- Harris, T. H., Banigan, E. J., Christian, D. A., Konradt, C., Wojno, E. D. T., Norose, K., et al. (2012). Generalized Lévy walks and the role of chemokines in migration of effector CD8+ T cells. *Nature* 486, 545–548. doi: 10.1038/nature11098
- Hilborn, R., and Mangel, M. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton NJ: Princeton University Press.
- Ho, D., Neumann, A., Perelson, A., Chen, W., Leonard, J., and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373, 123–126. doi: 10.1038/373123a0
- Hobbs, N. T., and Hilborn, R. (2006). Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecol. Appl.* 16, 5–19. doi: 10.1890/04-0645
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Stat. Sci.* 14, 382–401.
- Ioannidis, J. P. A., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C., et al. (2009). Repeatability of published microarray gene expression analyses. *Nat. Genet.* 41, 149–155. doi: 10.1038/ng.295
- Jewett, D. L. (2005). What's wrong with single hypotheses?: why it is time for strong-inference-PLUS. *Scientist* 19:10.
- Jin, X., Bauer, D. E., Tuttleton, S. E., Lewin, S., Gettie, A., Blanchard, J., et al. (1999). Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. *J. Exp. Med.* 189, 991–998. doi: 10.1084/jem.189.6.991
- Johnson, J. B., and Omland, K. S. (2004). Model selection in ecology and evolution. *Trends Ecol. Evol.* 19, 101–108. doi: 10.1016/j.tree.2003.10.013
- Kaech, S. M., and Cui, W. (2012). Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev. Immunol.* 12, 749–761. doi: 10.1038/nri3307
- Kaptchuk, T. J. (2003). Effect of interpretive bias on research evidence. *BMJ* 326, 1453–1455. doi: 10.1136/bmj.326.7404.1453
- Kirk, P. D. W., Babbie, A. C., and Stumpf, M. P. H. (2015). SYSTEMS BIOLOGY. Systems biology (un)certainities. *Science* 350, 386–388. doi: 10.1126/science.aac9505
- Kot, M. (2001). *Elements of Mathematical Ecology*. Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511608520
- Lofgren, E. T., Halloran, M. E., Rivers, C. M., Drake, J. M., Porco, T. C., Lewis, B., et al. (2014). Opinion: mathematical models: a key tool for outbreak response. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18095–18096. doi: 10.1073/pnas.1421551111
- MacCoun, R., and Perlmutter, S. (2015). Blind analysis: hide results to seek the truth. *Nature* 526, 187–189. doi: 10.1038/526187a
- Marino, S., and Kirschner, D. (2004). The human immune response to *Mycobacterium tuberculosis* in lung and lymph node. *J. Theor. Biol.* 227, 463–486. doi: 10.1016/j.jtbi.2003.11.023
- Mayer, J., Khairy, K., and Howard, J. (2010). Drawing an elephant with four complex parameters. *Am. J. Phys.* 78, 648–649. doi: 10.1119/1.3254017
- Meshkat, N., Eisenberg, M., and Distefano, III. J. J. (2009). An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner Bases. *Math. Biosci.* 222, 61–72. doi: 10.1016/j.mbs.2009.08.010
- Moher, D., and Altman, D. G. (2015). Four proposals to help improve the medical research literature. *PLoS Med.* 12:e1001864. doi: 10.1371/journal.pmed.1001864
- Mooney, D., and Swift, R. (1999). *A Course in Mathematical Modeling*. Mathematical Association of America.
- Noecker, C., Schaefer, K., Zaccheo, K., Yang, Y., Day, J., and Ganusov, V. V. (2015). Simple mathematical models do not accurately predict early SIV dynamics. *Viruses* 7, 1189–1217. doi: 10.3390/v7031189
- Nuzzo, R. (2015). How scientists fool themselves - and how they can stop. *Nature* 526, 182–185. doi: 10.1038/526182a
- O'Donohue, W., and Buchanan, J. A. (2001). The weaknesses of strong inference. *Behav. Philos.* 29, 1–20.
- Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* 263, 641–646. doi: 10.1126/science.263.5147.641
- Pabst, R., and Rosenberg, Y. J. (1998). Interpreting data on lymphocyte subsets in the blood of HIV patients - organ distribution, proliferation and migration kinetics are critical factors. *Pathobiology* 66, 117–122. doi: 10.1159/000028006
- Pandey, A., Atkins, K. E., Medlock, J., Wenzel, N., Townsend, J. P., Childs, J. E., et al. (2014). Strategies for containing Ebola in West Africa. *Science* 346, 991–995. doi: 10.1126/science.1260612
- Percus, J. (2012). *Mathematical Methods in Immunology, Courant Lecture Notes Series*. American Mathematical Soc.
- Platt, J. R. (1964). Strong inference: certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* 146, 347–353. doi: 10.1126/science.146.3642.347
- Popper, K. (2002). *The logic of scientific discovery*. New York, NY: Routledge Classics. Taylor & Francis.
- Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10:712. doi: 10.1038/nrd3439-c1
- Quinn, F. (2012). A revolution in mathematics? What really happened a century ago and why it matters today. *Notices AMS* 59, 31–37. doi: 10.1090/noti787
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., et al. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25, 1923–1929. doi: 10.1093/bioinformatics/btp358
- Scudellari, M. (2015). The science myths that will not die. *Nature* 528, 322–325. doi: 10.1038/528322a
- Segel, L. A. (1984). *Modeling Dynamic Phenomena in Molecular and Cellular Biology*. New York, NY: Cambridge University Press.
- Silberzahn, R., and Uhlmann, E. L. (2015). Crowdsourced research: many hands make tight work. *Nature* 526, 189–191. doi: 10.1038/526189a
- Simberloff, D. (2007). An angry indictment of mathematical modeling. *Bioscience* 57, 884–885.
- Steen, R. G., Casadevall, A., and Fang, F. C. (2013). Why has the number of scientific retractions increased? *PLoS ONE* 8:e68397. doi: 10.1371/journal.pone.0068397
- Vries, G. D., Hillen, T., Lewis, M., and Schönfisch, B. (2006). *A Course in Mathematical Biology: Quantitative Modeling with Mathematical and Computational (Monographs on Mathematical Modeling and Computation)*. SIAM. doi: 10.1137/1.9780898718256
- Wang, S., Hottz, P., Schechter, M., and Rong, L. (2015). Modeling the slow CD4+ T cell decline in HIV-infected individuals. *PLoS Comput. Biol.* 11:e1004665. doi: 10.1371/journal.pcbi.1004665
- Wei, X., Ghosh, S., Taylor, M., Johnson, V., Emini, E., Deutsch, P., et al. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373, 117–122. doi: 10.1038/373117a0
- Wodarz, D., May, R. M., and Nowak, M. A. (2000). The role of antigen-independent persistence of memory cytotoxic T lymphocytes. *Int. Immunol.* 12, 467–477. doi: 10.1093/intimm/12.4.467
- Wodarz, D., and Nowak, M. A. (2002). Mathematical models of HIV pathogenesis and treatment. *Bioessays* 24, 1178–1187. doi: 10.1002/bies.10196
- Zhang, Z., Tao, Y., and Li, Z. (2007). Factors affecting harellynx dynamics in the classic time series of the Hudson Bay Company, Canada. *Clim. Res.* 34, 83–89. doi: 10.3354/cr034083

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ganusov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.