Choosing the right model for policy decision-making: the case of smallpox epidemiology

Till Grüne-Yanoff¹

Received: 9 October 2017 / Accepted: 22 May 2018 / Published online: 30 May 2018
© The Author(s) 2018

Abstract Policymakers increasingly draw on scientific methods, including simulation modeling, to justify their decisions. For these purposes, scientists and policymakers face an extensive choice of modeling strategies. Discussing the example of smallpox epidemiology, this paper distinguishes three types of strategies: *Massive Simulation Models* (MSMs), *Abstract Simulation Models* (ASMs) and *Macro Equation Models* (MEMs). By analyzing some of the main smallpox epidemic models proposed in the last 20 years, it discusses how to justify strategy choice with reference to the core characteristics of these respective strategies. First, I argue that MEMs often suffice for policy purposes, and need to be replaced only if they are insufficiently robust. Such robustness results, however, only requires the use of ASMs, not MSMs. Second, I argue that although MSMs have larger potentials than ASMs in various dimensions, they are also more likely to fail—and that in many cases, this probability of failing outweighs their higher potential. In particular, these dimensions include the representation of the relevant target, the accurate measurement of the relevant parameters, the number of parameters included, the number of mechanisms modeled simultaneously, and the ways of dealing with structural uncertainty. While this in no way excludes the prospect that some MSMs provide good justifications for policy decisions, my arguments caution against a general preference for MSMs over ASMs for policy decision purposes in general and vaccination problems in particular.

Keywords Evidence-based policy · Epidemiology · Modeling · Macro versus micro · Microsimulations · Agent-based modeling · Robustness · Isolation

✉ Till Grüne-Yanoff
gryne@kth.se
<http://people.kth.se/~gryne/>

¹ Department of History and Philosophy of Technology, Royal Institute of Technology (KTH), Brinellvägen 32, 10044 Stockholm, Sweden

1 Introduction

Epidemic prevention and control are important functions of public health authorities. The *Centers for Disease Control and Prevention* (CDC), for example, has formulated contingency plans for outbreaks of Ebola, various types of influenza, smallpox and many other diseases. The CDC's response plan for smallpox, for example, prescribes a vaccination strategy that identifies and vaccinates contacts of an infected person.¹ Adoption of this strategy has not been uncontroversial, with some authors questioning the plan's efficacy (e.g. Bicknell 2002), recommending instead the voluntary mass vaccination of the population at the first indication of an infection.

Policymakers like the CDC want to make such policy decisions based on the best available scientific evidence. Yet smallpox has not occurred on an epidemic scale in modern urban societies, and therefore there is no empirical evidence about the differential efficacy of these and other vaccination strategies in e.g. a potential U.S. outbreak. Instead, policymakers and their advisors have resorted to modeling smallpox epidemics, and have used these models to investigate the efficacy of different vaccination strategies.

In this paper, I investigate how researchers chose the models on which vaccine policy decisions are supposed to be based. In particular, I will distinguish three types of models—*Macro-Equation Models* (MEMs), *Massive Simulation Models* (MSMs) and *Abstract Simulation Models* (ASMs). MEMs abstract the population into groups of different health status, and model disease dynamics between these groups. In contrast to this, MSMs and ASMs are micro-simulations, which model infection as the interactions of heterogeneous individuals. Recent technological advances have rapidly expanded the amount of detail that can be processed when computing a model. Micro-models whose detail is constrained largely by current computational capacities I call *Massive Simulation Models* (MSMs). Models whose detail is also constrained by other considerations (of simplicity, of transparency, etc.) I call *Abstract Simulation Models* (ASMs).² In each of these categories, numerous individual models have been developed and presented in the literature.

My main focus is methodological: I am interested in how model choice is justified. What justifies a modeler in preferring, say, a MSM to an ASM, or an ASM to a MEM? This justification obviously depends on the purpose for which the model is used. Smallpox modeling proves to be an excellent case here, because all models are employed for the same purpose—namely as evidence for the vaccine policy decision. However, my methodological interest goes beyond this case: by discussing the choice

¹ Smallpox was a highly infectious viral disease with a fatality rate of about 30% that was eradicated in all natural populations in 1980. Renewed interest in preparedness for a smallpox epidemic is based on scenarios involving a bioterrorism attack, either using a re-engineered virus, or using stock from stolen or rediscovered lab samples (Atkinson et al. 2005). For the CDC's response plan, see <https://www.cdc.gov/smallpox/bioterrorism-response-planning/public-health/vaccination-strategies.html>, last accessed 6.X.2017.

² Such constraints are widely discussed in the literature, for example Weisberg (2012, chapter 6). A model is transparent for a cognitive agent X at time t, if X at t knows all of the epistemically relevant steps involved in obtaining a model result from the model inputs. It is thus the opposite of epistemic opacity (Humphreys 2009, p. 618).

of models for smallpox, I hope to develop general arguments for model choice that can also be applied in other contexts.³

It might seem *prima facie* true that policymakers should prefer MSMs over ASMs or MEMs, because MSMs contain more detail, can be closer approximations to the real system, can better represent complexity and population heterogeneity, and the policymaker can use them as a holistic test-bed for potential policy interventions. MEMs and ASMs, due to their additional constraints, do not offer the same potential as MSMs in these regards—and therefore might be regarded as inferior for policy purposes.⁴

Against this *prima facie* claim, I argue that for policy purposes, ASMs or MEMs are often preferable over MSMs. I develop my argument in two steps. First, I show that policymakers for their purposes do not need the distinguishing properties of even the highest quality MSMs, if MEMs or AEMs are available. Secondly, I present a number of reasons why MSMs rarely live up to their potential. While MSMs potentially *could* be better models than MEMs or AEMs, they also run a greater risk of including errors, and therefore often *are not* better. Taken together, these arguments caution against relying on MSMs for policy purposes, while stressing the benefits that can be obtained from MEMs and AEMs.

The paper is structured as follows: Sect. 2 gives a conceptual distinction between MEMs, MSMs and ASMs and illustrates this difference with cases from the smallpox vaccination literature. Section 3 argues why MSMs are typically not needed if MEMs or AEMs are available. Section 4 discusses the problems with MSMs that make ASMs preferable for many policy purposes. Section 5 concludes.

2 Three types of models

Vaccination is one of the most effective ways of fighting epidemics. However, many vaccines do not provide long-term protection, or have serious side effects, so that a preventive vaccination (e.g., to all children aged five) is not feasible. Instead, these vaccines should be applied only when the risk of an epidemic is sufficiently high. The policymaker then has to make a momentous decision: namely, how to apply vaccinations in a large population when an epidemic is imminent or has already broken out. The most relevant alternatives are a *tracing vaccination* (TV), where the potential recent contacts of an infected individual are traced and vaccinated; a *limited vaccination* (LV), where a random subset of the population is vaccinated; or a *mass vaccination*

³ Similar discussions can be found for example in macroeconomics and in many engineering disciplines. In macroeconomics, the dominance of micro-founded DSGE models in policymaking is currently challenged both by those proposing agent-based simulation models (Alfi et al. 2009) and those proposing more traditional structural econometric models (Wren-Lewis 2018). Engineers for a long time have discussed the need to reduce the complexity of their models (Antoulas et al. 2001; Schilders et al. 2008), and more generally debate the advantages of analytic versus numerical modeling strategies.

⁴ Of course there is no principled reason why different types of models couldn't be pursued in combination. However, given limited resources in terms of attention, time and funding, scientists and policymakers *in practice* often face the exclusive choice between these model types. My discussion thus aims at those facing such a pragmatically driven, exclusive choice. Note however that at the end of Sect. 3 I develop an argument for the combined use of MEMs and ASMs in the special case of robustness analysis.

(MV), where the whole population is vaccinated. The choice is not trivial: MV is more likely to stop the spread of the disease, but is more costly and bears vaccination risks for a large population, while the effectiveness of LV and TV is less certain, but they are less costly, and do not expose a large number of people to vaccination risk. An instructive discussion in this regard is the evaluation of vaccine policies in the UK during the 2009 influenza pandemic, when an LV was implemented (Hine 2010). The case of smallpox is special, because smallpox vaccination is relatively safe and its long-term efficacy relatively high, even when administered several days after exposure (Atkinson et al. 2005). These particularities make TV a genuine alternative to either LV or MV for fighting potential smallpox epidemics.

Because smallpox has not occurred on an epidemic scale in modern urban societies, there is no empirical evidence about the differential efficacy of different vaccination strategies in e.g. a potential U.S. outbreak. Instead, policymakers and their advisors have resorted to *modeling* smallpox epidemics, and have used these models to investigate the efficacy of different vaccination strategies. Epidemiologists today face a bewildering array of models to choose from (for an overview, see Vynnycky and White 2010). In order to make this choice tractable, I will categorize these models into three groups: *Macro-Equation Models* (MEMs), *Massive Simulation Models* (MSMs) and *Abstract Simulation Models* (ASMs). Although epidemiological models can, of course, be distinguished further (or differently), this simple tripartite distinction will suffice for the present argument. I will describe these categories in turn, illustrate with an example and analyze their main distinguishing features.

MEMs abstract the population into *groups* of different health status, and model disease dynamics between these groups. For example, Kaplan et al. (2002) simulate an attack of 1000 initial smallpox cases on a population of 10 million. The population is assumed to mix homogeneously—i.e., its members have an equal chance of interacting with each other. R_0 , the rate of infections a single infectious agent generates among susceptibles, is assumed to be uniform throughout the simulation. $R_0 = 3$ is derived from historical data. An infected agent undergoes four stages. Only in the first is she vaccine-sensitive; only in the third and fourth is she infectious; in the fourth, however, she shows symptoms and is automatically isolated. Additionally, the administration of vaccinations is modeled under logistical constraints: MV of the whole population is achieved in 10 days. Tracking *and* vaccinating an infected person in TV, however, takes four times as many nurse-hours as a simple vaccination.

Kaplan et al. (2002) thus offer an example of a MEM. R_0 is a macro-parameter characterizing the population. Policy effects are modeled directly on this population parameter, and the main question is whether vaccine administration can outpace the random spread in the population. Unsurprisingly, perhaps, the results heavily favor MV over TV. Initiated on day five after the initial attack, MV leads to 560 deaths, while TV leads to 110,000 deaths. Sensitivity analysis shows that TV is more sensitive than MV to the size of initial attack and changes in R_0 , further supporting the strong results in favor of MV. The time to identify and then vaccinate the exposed is simply too long for the specified R_0 and for the assumed period in which the exposed are still sensitive to the vaccine.

Kaplan et al. (2002) belongs to the type of so-called compartmental models in epidemiology, in the tradition of Kermack and McKendrick (1927). Numerous variations

exist between these models, both with respect to the structural features and causal factors they do or do not include, as well as with respect to their parameterization.⁵ What connects them all is that they represent epidemics as a change in sizes of population subgroups (e.g. the “susceptible”, “infected” and “recovered” groups), where the change is governed by global parameters (e.g. the R_0). Such global parameters are attributes of the whole population, not of the individuals in the population. For this reason, I call these models *Macro-Equation Models* (MEMs).⁶

While MEMs are the most traditional type of epidemic models, they have recently been criticized for their macro-perspective. For example, the Kaplan et al. (2002) model has been criticized for its homogeneity assumption (Halloran et al. (2002). Critics argue that for smallpox infection, close and extended contact between infected and healthy agent is required. In a population of 10 million, they therefore consider it highly implausible that an infected agent has the same probability of having contact with any non-infected members. Furthermore, how the infected agents move through the population—i.e., with whom and with how many healthy agents they have contact—might influence the effects of different vaccination policies. Critics therefore claim that MEMs are not adequate models for vaccination policy decisions. Instead, they suggest that *micro-simulations* are required for this purpose.

Micro-simulations (also known as agent-based or individual-based models) model the interaction of individuals, including individual-to-individual infections, in order to account for macro-phenomena like an epidemic. Because of their individual-focused modeling approach, they have the potential to avoid the criticized homogeneity assumption mentioned above. For the same reason, micro-simulations are also computationally costly. Recent technological advances however have rapidly expanded the amount of detail that can be processed when computing a model result. Micro-models whose detail is constrained largely by current computational capacities I call *Massive Simulation Models* (MSMs). Models whose detail is also constrained by other considerations (of simplicity, of transparency, etc.) I call *Abstract Simulation Models* (ASMs). While the exact delimitation of these two categories necessarily is a gray zone, the two following examples will illustrate how different models from these categories really can be.

My example of an ASM, Burke et al. (2006), simulates a single initial infected person attack on a town network of either 6000 or 50,000 people. Town networks either consist of one town (uniform), a ring of six towns, or a ‘hub’ with four ‘spokes.’ Each town consists of households of up to seven persons, one workplace, and one school. All towns share a hospital. Each space is represented as a grid, so that each cell in the grid has eight neighbors. Agents are distinguished by type (child, health care worker [5% of adult population], commuter [10%], and non-commuter [90%])

⁵ Examples here include the choice between SIR (susceptible-infected-recovered) and SIS (susceptible-infected-susceptible) models, as well as between different sophistications of these basic models (e.g. SEIR, MSEIR or SEIS, cf. Vynnycky and White 2010). The Kaplan et al. 2002) model belongs to the SEIR group, as it includes an “exposed” compartment in addition to the standard SIR model.

⁶ These global parameter are typically represented as (differential) *equations*—hence the name. In principle every simulation model can, according to the Church Turing thesis, be expressed via equations involving global parameters (Epstein 2006). However, because it is the local and heterogeneous properties of individual agents that are assumed to drive ASMs and MSMs, such a representation is almost never chosen in practice.

by family ID and by infectious status. Each ‘day,’ agents visit spaces according to their type, and then return home. On the first ‘day’ of the simulation, the position in schools and workplaces is randomly assigned, but after that, agents remember their positions. During the ‘day,’ agents interact with all of their immediate neighbors: 10 times at home, 7 times at work, and 15 times in the hospital. After each interaction, they move positions to the first free cell in their neighborhood. Homogeneous mixing is thus completely eschewed; instead, agents interact in a number of dynamic neighborhoods.

Transmission occurs at a certain rate in each of the agents’ interactions. It can infect both contactor and contacted. Transmission rates depend on the stage the infectious person is in, the type of disease he has, and whether the susceptible agent has partial immunity.

Burke et al. assessed only TV as a first policy intervention, and LVs of varying degrees only as ‘add-on’ measures. Results for all three networks show substantial concordance. Contrasted with a ‘no response’ scenario, TV in combination with hospital isolation was sufficient to limit the epidemic to a mean of fewer than 48 infections and a mean duration of less than 77 days. Post-release LV of either 40 or 80% of the total population added some additional protection, reducing the mean of infected people to 33 and shortening the mean duration to less than 60 days.

My example of an MSM, Eubank et al. (2004), simulates an attack of 1000 infected agents on the population of Portland, OR, of 1.5 million. Portland is represented by approximately 181,000 locations, each associated with a specific activity, like work, shopping, or school, as well as maximal occupancies. Each agent is characterized by a list of the entrance and exit times into and from a location for all locations that person visited during the day. This huge database was developed by the traffic simulation tool TRANSIMS, which in turn is based on US census data.

Smallpox is modeled by a single parameter, disease ‘load’ (analogous to a viral titer). Agents have individual thresholds, above which their load leads to infection, symptoms, infectiousness, and even to possible death. Every hour, infectious agents shed a fixed fraction of their load to the local environment. Locations thus get contaminated with load, which is distributed equally among those present. Shedding and absorption fractions differ individually. Infected individuals withdraw to their homes 24 h after becoming infectious.

The Eubank et al. (2004) model yields roughly similar results for both vaccination policies. MV with a 4-day delay resulted in 0.39 deaths per initially infected person; TV with the same delay in 0.54 deaths. Varying delays, they found that delay in response is the most important factor in limiting deaths, yielding similar results for TV and MV.

Both ASM and MSM seem to yield similar policy recommendations. To quote just two examples: “[C]ontact tracing and vaccination of household, workplace and school contacts, along with effective isolation of diagnosed cases, can control epidemics of smallpox” (Burke et al. 2006, p. 1148); and “[O]utbreaks can be contained by a strategy of targeted vaccination combined with early detection without resorting to mass vaccination of a population” (Eubank et al. 2004, p. 180).

Staring from these examples, what are the main differences between MSMs and ASMs? MSMs differ from ASMs at first glance by their much higher level of detail, especially the number of variables and parameters they include, and the number of

relation between these. The Eubank et al. (2004) model, for example, includes approximately 1.6 million vertices with maximally 1.5 million edges that might change 24 times a day; while the Burke et al. (2006) model includes only approximately 7900 vertices with maximally 6000 edges (and approximately 55,000 vertices for the larger one) that can change maximally 17 times a day.

Based on the richness in realistic detail, MSMs are sometimes claimed to offer a highly accurate picture of the real system:

Such models allow for the creating of a kind of virtual universe, in which many players can act in complex — and realistic — ways. (Farmer and Foley 2009, p. 686)⁷

Understood in this way, MSMs are typically seen as direct representation of real systems: their structure is designed to allow a mapping from the model to the target without taking recourse to mediating models.⁸ The Eubank et al. (2004) model, for example, is introduced as a direct representation of the city of Portland. ASMs, in contrast, can hardly ever claim to represent a real system directly—their level of detail is not sufficient. At best, they are able to represent *stylized facts* about or abstractions of a system, which have been prepared through an abstraction or idealization procedure from the real system. The Burke et al. (2006) model, for example, explicitly claims to represent an “artificial city” that shares some properties with real cities, but is different otherwise (Burke et al. 2006, p. 1142).

That MSMs represent real systems directly is further supported by the practice of fitting or calibrating the model directly to real data. The Eubank et al. (2004) model, for example, bases the specification of edges and their hourly change on census data. The Burke et al. (2006) model, in contrast, stipulate certain edge changes and interpret these activities as “being at home,” “going to work,” “being at the hospital,” etc. These interpretations are not based on data from actual target systems, but rather rely on plausibility intuitions. Alternatively, where ASMs are calibrated to data, this design process is not direct: it is based on interceding models, which abstract stylized facts of “going to work” or “being at the hospital” from available data, and then allow the representation of these abstractions in the ASM.

Both MSMs and ASMs typically represent processes of their target systems. But even here, there is an important difference. MSMs typically represent a multitude of simultaneous processes or mechanisms, while ASMs typically represent only one or a small number of such processes. Presumably, many of these components will operate through different mechanisms. Consequently, putting them all together in a single MSM implies that many different processes will operate simultaneously when producing a model outcome. The Eubank et al. (2004) model, for example, distin-

⁷ This and the following two quotes are by economists writing about economic MSMs, not specifically about the epidemiological MSMs I discussed here. However, debates in economics about the macro–micro divide, about the precision and amount of detail in modeling closely parallel the issues I am addressing here. I am quoting these authors to illustrate the motivation for choosing MSMs generally, which *mutatis mutandis* applies to the case under consideration.

⁸ By characterizing MSMs as direct representations of their token targets, I am referring merely to the modelers’ (implicit or sometimes explicit) intention that these models *should* allow such a mapping, not that they actually satisfy the requirements of such strong representational relationships.

guishes several activities at each location, each of which yields different contact rates; it also includes the effects of demographic factors (age in particular) on mixing; it distinguishes different forms of smallpox; and it tries to incorporate at least some rudimentary effects of infection on behavior. The Burke et al. (2006) model, in contrast, includes a lesser number of locations and did not distinguish activities or demographics, nor does it include infection effects on behavior. Thus, MSMs typically include many more simultaneous mechanisms than ASMs.

So far, I have distinguished MSMs and ASMs only with respect to their different representational relations to a target. Additionally, I also distinguish them with respect to how these relations are interpreted and for what purposes they are used. Regarding interpretation, the greater amount of detail in their models is typically employed by MSM modelers in order to achieve a “realistic” interpretation of the model’s representational function. “Realisticness” is a subjective psychological effect that might stem from an impression of familiarity and might lead to greater trust in the model and its conclusions:

[D]ecision makers might be more willing to trust findings based on rather detailed simulation models where they see a lot of economic structure they are familiar with than in general insights obtained in rather abstract mathematical models. (Dawid and Fagiolo 2008, p. 354)

In the Eubank et al. (2004) model, for example, the user can follow the development of the epidemic on a map of Portland, OR. The map offers a high-resolution representation of residential density, and as the epidemic progresses, the proportion of the population infected at one node is shown, as well as aggregate numbers of fatalities, vaccinations and quarantines (if applicable). The user thus gets a realistic impression of a possible epidemic development in the city. ASMs do not offer such a rich collection of familiar details, are therefore typically not considered as “realistic” as MSMs, and thus might not inspire the same amount of trust.

The features discussed so far also imply an important difference in the use of these model types for policy decision-making. MSMs are often used as part of a “holistic approach”: a “model of the whole economy” (Farmer and Foley 2009, p. 686) is used as a “virtual universe” (ibid.) to evaluate the effects of proposed interventions in the target system. That is, interventions are simulated in the model, and model results are interpreted as forecasts of the results of such interventions in the real system. While Eubank et al. (2004) do not explicitly endorse such a use of their model, it is at least compatible with it. ASMs cannot be used in this way, as they do not offer a representation of the whole system or of the combination of its many operating mechanisms. The holistic synthesis that the MSM promises as part of its package must be performed by the ASM user in some other way, for example through expert judgment.

To be clear, MSMs and ASMs have important similarities, despite the differences I discussed. Both aim to represent non-linear and complex behavior, albeit on different levels of abstraction and idealization—it is through this that they distinguish themselves from MEMs. Furthermore, both types of models abstract and idealize, but to different degrees, and for different reasons. MSMs abstract and idealize for tractability and computation reasons: they are mainly constrained by current computational capac-

ities. ASMs, in contrast, are also constrained by other considerations (for example, of simplicity, of transparency, etc.), so that computability rarely becomes a relevant constraint for them.

Last, the distinction between MSM and ASM is itself a simplification. Many actual simulation models exhibit some properties of the one kind and some of the other, and thus do not clearly fall into either category. However, this does not pose a problem for my argument; my discussion in Sect. 4 addresses these properties separately so that relevant conclusions can be drawn for such “in-between models,” too. While I am aware of the possibility of such cases, I have nevertheless decided to stick with the dichotomous distinction for ease of exposition.

3 MSMs not necessary for policy decisions

Epidemiologists face a methodological choice: they need to select the model they want to use from a large menu of possible models, and they have to justify why this is the best choice for their purposes. In real practice, of course, this menu will consist of a large and unwieldy set of possibilities. But for the purposes of my argument, it suffices to discuss this methodological problem as a choice between three modeling strategies—MEMs, ASMs and MSMs. My aim is to present and analyze the reasons for choosing one option over another, for the purpose of policy decision-making.

In this section, I argue that policymakers for their purposes do not need the distinguishing properties of even the highest quality MSMs, if MEMs or AEMs are available. Defenders of MSMs might argue that they can be used as part of a “holistic approach”: the model is claimed to be a faithful copy of the real system of interest, representing *all* factors relevant for the target variable, so that any intervention simulated in the model is interpreted as a forecast of the *total* results of such interventions in the real system. Such a view obviously has a lot of appeal, not least to policymakers. Whether MSMs can deliver such a promise, I will discuss in Sect. 4. But first, I argue that such a holistic approach is not the only way how models can support policy decision-making, thus explaining why MSMs are not necessary for such purposes.

The basic idea is that a decision maker can successfully perform interventions based on the knowledge of *one* mechanism, even in the absence of knowledge of other causes present in such a system. This knowledge can be represented by isolating models. Such isolating models disregard the operation of other factors, even if in reality these factors are often or always present.

Consider for example a captain steering a boat through a rough sea. To maintain her ship on a save course, she does not need to predict all the factors that affect the boat’s course in advance, nor does she need to predict the interaction of these factors with her controlling intervention. Rather, to control her boat it is enough that she knows the operation of the steerage—e.g. that a quarter turn on the ship’s wheel translates into a 5-degree shift of the rudder, and that this rudder shift has an effect on the boat’s course. The captain might have acquired this knowledge by studying an isolating model of the steerage. Other factors (e.g. wind, current, obstacles) obviously also influence the boat’s course. As these factors are not included in an isolating model, such a model does not provide the *total* effect of the intervention. Yet it is enough that the captain

employs her knowledge of the steering mechanism, and that she correctly observes how other present factors affect the course, in order to maintain control of the boat.

Similarly with the public health official steering through an epidemic. As long as she knows reliable mechanisms through which she can influence the course of the epidemic, she does not need to be able to predict the development of the whole system, nor the interaction of other parts of the system with her intervention. Instead, the application of this knowledge would depend on observed conditions of the state of the epidemic and how it is developing, as expressed here:

“Would an attack be small and controllable through traced vaccination or large enough to require mass vaccination? Would an attack be overt, in which case it could prove possible to respond immediately in a highly targeted fashion and obtain much better results, or covert and detected only from symptomatic cases as assumed in this article?” (Kaplan 2004, p. 269)

Kaplan can ask these precise questions about the epidemic to be controlled because in his models, he *isolated* a purportedly reliable mechanism for controlling “small” epidemics, another one for “large enough” ones, one for overt attacks and another for covert ones. Such knowledge allows policymakers to react to the conditions of an actual epidemic, by choosing the appropriate mechanism that allows her to influence the system in the desired direction.

Of course, my above argument depends on the questions (i) whether there is such an isolatable mechanism, and (ii) that the model—in this case the MEM—accurately represents this mechanism. Answers to both of these questions require empirical evidence—how this is obtained is beyond the scope of this paper (for a discussion of the importance of mechanistic evidence for policy making, see Grüne-Yanoff (2016a)). This holds in particular for the case of smallpox, where, as mentioned before, no scientifically documented epidemics exist. In the absence of such evidence, models merely represent *possible mechanisms*. Yet not all models are equally good candidates: some are better candidates for isolating mechanisms than others. In particular, robustness with respect to idealizations of the isolated mechanism provides such an assessment criterion. To this I turn now. The concept of isolation in modeling has been extensively analyzed in the literature (e.g. Mäki 1992; Cartwright 1994). A causal relation between an intervention and an effect is isolated by excluding the impact of some ancillary factors from the model. This can be achieved either by omitting these factors from the representation completely, or by idealizing them (i.e. changing a parameter in the representation to a different value, typically to zero or infinity).

While idealization thus distorts, one can distinguish the isolation process, which may involve idealization, from the product of isolation. When isolating a factor F from intervening factors $G_1, \dots, G_n$, one may either omit or idealize the operations of the G_i , but not the operation of the factor F itself. This way, one makes false claims about the G_i s; but the purpose of the theoretical process—to isolate the operation of F —remains intact. Idealization thus is a procedure applied to entities one isolates from, but not to entities that one intends to isolate (Mäki 1992, p. 328); and idealization is used as an auxiliary technique for generating isolation, yet it is not part of isolation itself (Mäki 1992, p. 325). Thus the product of isolation—the isolated factor—must never be idealized.

Often isolating models do not only idealize factors outside of the isolated mechanism, but they idealize the isolated mechanism itself. This is in particular true for MEM, due to their highly abstract nature. Kaplan’s model, for example, includes many idealizing assumptions concerning the isolated mechanism, including “all random tracing is of susceptibles” or “all people leave the untraced compartment via tracing, not disease symptoms” (Kaplan et al. 2003, p. 46) that are clearly assumptions about the isolated mechanism itself. This poses a potential problem for the accurate representation of the relevant mechanism by such isolating models, and thus for their use in policy decisions as described above. Idealizing aspects of the isolated mechanism itself threatens the reliability of the model result, also for policy purposes (Cartwright 2007; Grüne-Yanoff 2011a).

To continue the analogy, if the captain is given a highly idealizing representation of the steering mechanism—e.g. where the wheel is connected to the rudder through a rigid rod, rather than tiller ropes –, she might legitimately doubt that the information provided by this model is relevant for her successfully steering the ship.

A way how to alleviate her doubts is to show that these idealizations of the isolated mechanism itself are innocuous—i.e. although they are not accurate of the represented target, their particular form makes no difference to the model result. By varying the assumption in question, and observing that the model result does not change under such variations, a modeler shows that the model is *derivationally robust* under the assumption in question. Robustness results justify the modeler to be confident that the assumption, although idealizing part of the mechanism, does not affect the model’s result (Muldoon 2007; Kuorikoski et al. 2010; Lloyd 2010). If that is the case, then there is little reason to fear that the model’s reliability is compromised by this assumption. Note however that this concerns *only* the robustness of idealizations of the isolated mechanism itself. As I discussed above, any isolating models idealizes ancillary factors, and to the extent that these factors matter at all, one should not expect that the isolated mechanism is robust with respect to them.

Robustness plays an important role in the debate around homogeneous mixing in smallpox modeling. As mentioned above, Halloran et al. (2002) criticize the Kaplan et al. (2002) MEM for neglecting the effects of non-random network structure on the difference in effectiveness between MV and TV. In Kaplan’s et al. MEM, MV does better than TV—yet when Halloran et al. relaxed the homogeneity assumption, and modeled a more structured interaction between individuals in their ASM, then TV did better than MV. Kaplan and Wein (2003) responded by showing that both models yield roughly the same results (MV more effective than TV) if both models are compared on the same scale. Halloran and Longini (2003) replied that although this might be true for small populations and low number of initial infections, it is unlikely to be the case for larger populations and many initial infections.⁹

This debate illustrates how investigations of model robustness proceed. Clearly, Halloran et al. (2002) had some *prima facie* reasons to believe that the homogene-

⁹ They did not model, however, such a scaled up version. Instead, they only point to the possibility: “Many choices are available in scaling up the heterogeneous [ASM] model.... Differences between the two current models [Kaplan’s et al. MEM and their ASM] could also become more manifest when scaled up” (Halloran et al. (2002, p. 1503)).

ity idealization in the isolated mechanism made the model less reliable for deciding between MV and TV. They therefore constructed an ASM that relaxed that assumptions and sought to show that indeed it does make a difference for the policy decision at hand. However, Kaplan and Wein (2003) could show that it wasn't the idealization that made a difference it rather was the scale. The MEM thus was robust with respect to the homogeneity assumption, at least at a small scale (i.e. small populations and low number of initial infections). Nevertheless, it remained unclear whether the MEM would also be robust at a larger scale. Halloran and Longini (2003) do not offer a result that shows that it isn't, but give various reasons to think why it might not.

Without adjudicating this debate, it is clear that the arguments from both sides rely on the use of *both* MEMs and ASMs. Halloran et al. (2002) needed the ASM to relax the idealization in question. But Kaplan and Wein (2003) also needed the ASM in order to show that this relaxation did not make a difference. Halloran and Longini (2003), furthermore, would need a modified ASM (from the “many choices ... available to scale up”) in order to support their claim that on larger scale, the MEM is not robust. Robustness requires model comparison across different types of models—and in epidemiology, in particular, between MEMs and ASMs. Thus, even if in the end everybody agreed to accept the MEM as the relevant model, the ASMs would have served the important evidential functions in this acceptance decision.¹⁰

The MSM, however, is *not* needed for this purpose. Why not? Because the model comparison relevant for robustness questions always is one of relaxing one or a few specific assumptions. To do this, one does not need massive detail, multiple mechanisms, direct representation of a real target, or calibration with real data. In the epidemiology case, it required switching from a macro- to a micro-simulation. But the micro-simulation could still be an isolating model and thus be highly simplified in many respects. ASMs are often sufficient for this. Thus, the policymaker typically will not need an MSM to serve her purposes.¹¹

4 MSMs more likely to include errors than ASMs

While MSMs might not be necessary to arrive at reliable evidence for policy-making, it would of course be convenient for a policy maker to have a model that predicts and represents the accurate course of an epidemic in advance, including all environmental influences, and correctly predicting the effect of all possible interventions.

To this I reply that MSMs rarely satisfy the conditions under which such convenient use would be justified. Essentially, my argument is that, although MSMs have larger potentials than ASMs in various dimensions, they are also more likely to fail—that is, model users are more likely to make erroneous inferences with them. In many cases,

¹⁰ For a similar argument, also in the context of isolating models, see Gräbner et al. (2017). For the domain of economics, they argue that agent-based and equation-based approaches should be considered complements rather than substitutes.

¹¹ The qualifiers ‘often’ and ‘typically’ are important here, though. If serious attempts to isolate relevant mechanisms fail—e.g. because all factors influencing the effect are strongly interdependent – then attempts to represent the interaction of all factors would naturally lead to a form of MSM. Reasons for doubting that this is a viable option I discuss in the next section.

this probability of failing outweighs their higher potential. Making use of the above smallpox models, I now discuss the different dimensions in which MSMs might fail in comparison to ASMs, and why therefore the latter might be preferable to the former for policy purposes.¹² These dimensions are conceptually separate, although in practice they often overlap.

4.1 What is the target?

Prima facie, MSMs like Eubank et al. (2004) have a particular (token) target: for example, the town of Portland, OR. ASMs like Burke et al. (2006), in contrast, do not appear to have such a particular target; rather, they represent an abstracted type, like “a town” or “an urban population network.” Consequently, MSMs are often judged to be more *realistic* than ASMs, as model users can more easily trace the MSM features to the properties of a particular target. This realisticness judgment, in turn, as the above quote from Dawid and Fagiolo (2008) shows, often induces policymakers to place more trust in the reliability and usefulness of the model in question. For this reason, MSMs often seem preferable to ASMs for policy purposes.

But is the inference from realisticness to reliability and usefulness justified? Presumably, the argument is that (i) judging a model to be realistic indicates that it is a highly accurate representation of the target, and that (ii) a highly accurate representation of the target is a sufficient condition for the model to give reliable and useful information about possible policy interventions in the target.

While I do not dispute these claims individually here, I argue that their conjunction does not constitute a valid argument if the meaning of “target” changes between them. This is precisely what happens in the smallpox simulation studies. The target *of the policy question* is the city environment generally, as the introductory sentence of Eubank et al. (2004, p. 180) shows:

The dense social-contact networks characteristic of urban areas form a perfect fabric for fast, uncontrolled disease propagation. [...] How can an outbreak be contained before it becomes an epidemic, and what disease surveillance strategies should be implemented?

Furthermore, because epidemic policies are typically the responsibility of national or international institutions, the targets of the policy question are all cities within the governing domain of that institution (e.g., all US cities, all cities in industrialized countries, all cities of the world, etc.). The target of such a policy question thus is an abstract entity: the network characteristics of all urban areas within the relevant domain.

The *model’s target* in the MSM case, in contrast, is a particular target: the city of Portland, OR. The authors of this model suggest that it is just an instance of the network

¹² Earlier discussions of some of the dimensions in which MSMs and ASMs differ can be found in Grüne-Yanoff (2011b, bb).

characteristics in urban areas.¹³ But by choosing a particular target, they allow for a possible divergence between the meaning of “target” in step (i) and (ii) in the above argument. In particular, the judgment that their model is realistic might now be based on relational features of their model and the city of Portland that are wholly irrelevant for the relational features of their model and network characteristics of all urban areas within the relevant domain. For example, inclusion of the Columbia riverbed, of the locations of Portland’s universities, as well as Portland’s public transport system, might increase the realism of the model. However, these might be features that are either *irrelevant* for the path of an epidemic through an urban network, or they might not be representative of urban networks in the US more generally. Both of these cases might sever the relation between realism, reliability, and usefulness: an MSM with these features might be more realistic than a ASM, while the ASM is a more accurate representation of the general network characteristics of all urban areas within the relevant domain. In such cases, the ASM would be a more powerful policy tool than the MSM.

4.2 How to measure parameters

MSMs differ from ASMs in their much higher level of detail, especially the number of variables and parameters they include, and the number of relations between these. Assuming that both models have the same target (so that problem 4.1 does not arise), a higher number of variables and parameters gives MSMs more potential than ASMs to accurately represent the target system. *Prima facie*, this gives MSMs an advantage over ASMs for policy purposes.

However, this argument assumes that the additional variables and parameters that give MSMs an advantage over ASMs can be measured or estimated with sufficient accuracy. Both of these assumptions are problematic. I will discuss measurement problems in this subsection and estimation problems in the next.

The measured variables and parameters of the smallpox MSM are those whose value is directly obtained from some external data source. For example, properties like age, occupation, health, and home location are obtained from census data for all of the 1.5 million agents in the model. Properties of the urban transport network and of land occupation and use are obtained from urban planning organizations (Eubanks et al. 2004, Supplement, p. 3). These examples of massive data intake seem indeed to support the comparative detail richness of MSMs over ASMs.

However, a closer reading of the article and its supplementary material reveals that many of the parameters and variables could not be accurately measured (or even measured at all). Instead, they are determined by ad hoc assumptions, best guesses, or the use of reasonable ranges. I describe three instances here for illustrative purposes. The first concerns the disease-relevant contacts of agents within a location, which cannot be found in census data:

¹³ “We view the social networks created by TRANSIMS as a single instance of a stochastic process defined in an enormous space of possibilities” (Eubank et al. 2004, Supplement, p. 3).

We do not have data for proximity of people, other than that they are in the same (possibly very large) location. [...] It seems as though the dependence on distance is very coarse: one mode of transmission occurs at close ranges (<6 feet) and another for large ranges. We have developed an ad hoc model that takes advantage of this coarseness. (Eubanks et al. 2004, Supplement, p. 9)

This ad hoc model makes uniform assumptions about the occupancy rate of locations within a city block that are, the authors admit, “nothing more than reasonable guesses” (Eubank et al. 2004, Supplement, p. 11). Location occupancy rates, however, crucially influence the number of possible contacts—and hence may be relevant for the spread of disease.

Another example concerns the parameterization of the disease model:

There is not yet a consensus model of smallpox. We have designed a model that captures many features on which there is widespread agreement and allow us to vary poorly understood properties through reasonable ranges. (Eubanks et al. 2004, p. 183)

What “reasonable” means in this context, and how much it is related to available data, remains unclear. Finally, here is an example concerning the parameterization of the TV intervention:

Every simulated day, if contact tracing is in effect, a subset of the people on the list [of people showing symptoms] is chosen for contact tracing. [...] In the experiment reported here, we use the fraction 0.8 and set the absolute threshold at either 10,000 or 1,000. These are probably unrealistic numbers, but they allow us to estimate the best case results of a targeted vaccination strategy. (Eubanks et al. 2004, Supplement, p. 11)

In all of these examples, the very detail-demanding parameter and variable set poses the question of how to fill them with content. By default, one might assume that they are filled with empirical data. But it turns out that for these examples, empirical data are not available, or of too low a quality. So the modelers instead resorted to ad hoc assumptions, best guesses, or reasonable ranges. This is not an idiosyncrasy of the models discussed here, but follows directly from the definitions of ASMs and MSMs: MSMs contain many more parameters and variables than ASMs, therefore impose more demanding requirements on measurement, which often cannot be fully met. While MSMs of course also contain many idealizations, these differ from the ad hoc assumptions discussed here in two ways. First, in ASMs these assumptions are explicitly identified as idealizing assumptions, while in MSMs they often appear as being the result of observations. Second, because of ASMs’s more abstract nature, the number of such assumptions is typically lower than in MSMs, making them easier to recognize and control.

I do not intend these observations as criticisms of the particular smallpox model, or of MSMs more generally. It seems perfectly reasonable to improvise on some parameters of one’s model. But when discussing model choice, and in particular how to choose the resolution of detail of one’s model, one should be mindful how this choice affects the need to improvise. Imagine an extreme case, where a detail-poor

model with only a few parameters that all can be determined from high-quality data can be developed into a detail-rich model, whose parameters can be filled only by ad hoc assumptions, best guesses, or use of reasonable ranges. Because these improvisations carry a large chance of error, the detail-poor model is likely more accurate and therefore preferable for policy purposes than the detail-rich model. My MSM vs. ASM case is much less clear-cut than this extreme case, firstly because the parameters of the ASM are typically determined in a haphazard way, too, and secondly because the MSM does include a lot of certified data. However, there is a similar trade-off as in the extreme case, and that trade-off might in some cases lead to the conclusion that the ASM is a more powerful policy tool than the MSM.

4.3 Number of parameters

Assume that the measurement of parameters was not a problem, so that Sect. 4.2 would not impose any constraints on the amount of detail incorporated in an MSM. In that case, another argument against such unchecked increase of detail arises from the comparative performance of such models in parameter estimation or calibration.

Disregarding technical detail, estimation and calibration both aim to determine values of unobservable model parameters by fitting the model to observable data. In the smallpox case, many parameters of the underlying TRANSIMS and EpiSims models are thus determined. To put it simply, the model takes census data, transport network data, land use data, etc. as inputs, and gives as output contact incidences, duration, and locations between individual agents. In accord with the generative program in simulation studies (Epstein 1999), model parameters are then adjusted so as to generate that model result that fits best with observational data. Once a close enough fit to such data has been achieved, the model is considered validated, and counterfactual policy interventions are introduced.

At first sight, MSMs appear to be better equipped to perform well in estimation or calibration exercises. If the target is of high complexity (which, in the case of vaccination policies, it undoubtedly is), then the more constraints one imposes on the model (in terms of the nature and number of its parameters), the less well such a model can fit the target. Conversely, the fewer constraints are imposed on a model, the better it can fit its target. Thus, it seems that MSMs can achieve a better fit to their targets than ASMs, and therefore appears as the more powerful policy tool.

The above intuition, although correct, misses an important trade-off that is well known in the model-selection literature. Although models with more free parameters have a larger potential to fit the target well, the larger number of free parameters often yields a lesser fit than the one achieved by a model with fewer parameters.

This trade-off becomes clearer by distinguishing two steps in the process of fitting a model to data. The first step consists in selecting a model—i.e., in specifying the number of parameters. Here, increasing the number of parameters indeed increases the model's *potential* to accurately represent the target.

The second step consists in calibrating or estimating the parameters based on a data *sample* drawn from the population. Increasing the number of parameters increases the model's fit to the sample—but this is not the ultimate goal. Rather, increasing

the model's fit *to the target* is. Fitting the model “too closely” (i.e., by including too many parameters) to the sample will pick up on the inevitable random error in the sample, and thus leads to an increase in the divergence between model and target. This phenomenon is well known as “overfitting” in the statistics and machine-learning literature, and it applies to simulation modeling as well (Myung 2000).¹⁴

Selecting the right number of free parameters thus is the problem of “finding an appropriate compromise between these two opposing properties, *potential* and *propensity to underperform*” (Zucchini 2000, p. 45). As various studies have shown, if the sample size is large, adding more parameters above a certain threshold will not substantially increase fit to target; if sample size is medium or small, adding more parameters even decreases fit to target (Zucchini 2000; Gigerenzer and Brighton 2009).

This general finding also applies to the choice between MSM and ASM. In Sect. 2, I defined MSMs as containing many more parameters than ASMs. Consequently, MSMs are more subject to the danger of overfitting, and therefore more likely to fit the underlying target badly. Of course, whether in a particular case of comparing an MSM and an ASM the trade-off will favor one or the other is an open question (in particular, this is also the case for the two smallpox models, as a numerical study of their respective fit is beyond the scope of this paper). However, this general tendency makes it implausible to generally prefer MSMs over ASMs for policy purposes.

4.4 Number of mechanisms

One of the important features of the simulation models discussed here is that they explicitly aim to represent *mechanisms*. In the smallpox case, both the MSM and the ASM were introduced as improvements over Kaplan's et al. (2002) macro model, because they explicitly modeled the population mixing mechanism instead of simply assuming homogeneous mixing. Nevertheless, the MSM and the ASM differ substantially in how they introduce such additional mechanisms. The smallpox ASM seeks to introduce a small number of simple mechanisms, while the MSM introduces a multitude of detail-rich mechanisms that are assumed to operate simultaneously.

In particular, the MSM distinguishes several activities at each location, each of which yields different contact rates; it also includes the effects of demographic factors (age in particular) on mixing; it distinguishes different forms of smallpox; and it tries to incorporate at least some rudimentary effects of infection on behavior. The ASM, in contrast, includes a lesser number of locations and does not distinguish activities or demographics; nor does it include infection effects on behavior.

Most observers seem to see the inclusion of additional mechanisms in comparison to the Kaplan et al. (2002) model as beneficial. It then also seems *prima facie* plausible to prefer the MSM to the ASM, as the former includes even more mechanisms and mechanistic detail than the latter.

¹⁴ This issue further compounds the problem of particular model targets when policy targets are more abstract, discussed in Sect. 4.1. A close fit to the particular *model target*—even without the problem of overfitting—might not improve the model's usefulness for questions about the abstract *policy target*.

Countering this intuition, I will use an argument made against the purported higher explanatory power of realistic simulation models. This argument has been put forward by Lenhard and Winsberg (2010), amongst others, with a specific focus on climate models. In short, they argue that with increasing complexity, models get more and more *opaque*; and this opacity prevents or at least reduces understanding the model components' contributions towards the model outcome.

More specifically, Lenhard and Winsberg argue that, with increasing complexity, the “fuzzy modularity” of a model increases. The more complex a model, the more subcomponents it has. Furthermore, when running a simulation on a complex model, these model components are run together and in parallel. But they do not all independently contribute to the model result. Rather, the components, in the course of a simulation, often exchange results of intermediary calculations among one another—so that the contribution of each component to the model result in turn is influenced by all those components that interacted with it.

The results of these modules are not first gathered independently and then only after that synthesized. [...] The overall dynamics of one global climate model is the complex result of the interaction of the modules — not the interaction of the results of the modules [...] Due to interactivity, modularity does not break down a complex system into separately manageable pieces. (Lenhard and Winsberg 2010, p. 258)

To put it differently, the effect of the multiple mechanisms is underdetermined more in an MSM than an ASM: first, due to the larger number of mechanisms included in an MSM, but also due to the increased interaction—the “fuzzy modularity”—of the mechanisms in the MSM. Clearly, there is more fuzzy modularity in a MSM like Eubanks et al. (2004) than in an ASM like Burke et al. (2006). In the first place, this is a problem for the explanatory power of MSMs. Although MSMs might generate the *explanandum* quite closely, because of the higher degree of underdetermination, it is more difficult in MSMs than in ASMs to infer from this fit which of the modeled mechanisms contributed to the generated result. If understanding consists in identifying the mechanisms that produced the *explanandum*, then a model's fuzzy modularity undermines improvements in our understanding.

Now, policy makers do not need to worry about whether the models increase understanding. So why should fuzzy modularity be a problem for them? Due to fuzzy modularity, MSM users do not know *how* individual mechanisms contribute to the production of a relevant effect. But knowing how individual mechanisms contribute is highly relevant both for (i) designing and for (ii) justifying interventions. First, without knowing how individual mechanisms contribute, the designer does not know *where* to intervene, as intervention in one contributing cause might yield multiple effects—through multiple mechanisms—that might amplify each other or cancel each other out. Furthermore, we don't know whether the intervention—effect relation can be transferred to other contexts, where some of the parallel mechanisms might be operating differently. Something like this is the case in Eubanks et al. (2004): their results might depend on some or all of the mechanisms in the model, or on their specific interaction, but it is impossible for the modelers to pry these influences apart.

Second, without knowing how individual mechanisms contribute to a result, we do not exactly understand how and why the results emerge. In the extreme case, an epistemically opaque model does not provide an answer why an intervention X leads to a desired effect, while an intervention Y does not. Thus, such a model fails to provide *justifying reasons* for an intervention. Instead, the modeler is reduced to argue that one should adopt X rather than Y because the model shows X to be better than Y, which does not constitute a justifying reason at all—as long as the modeler has not argued why one should trust the model. But trusting the model requires identifying the factors that produce the model result, which is exactly what epistemic opacity excludes. Thus, because ASMs are generally less epistemically opaque than MSMs, they might be preferable for policy purposes.

4.5 Structural uncertainty

From the discussion so far (as well as from common sense), it follows that uncertainty in model specification can never be fully eliminated, however little or much detail one might want to include in one's model. Some sources of uncertainty affect MSMs more than ASMs, as discussed in Sects. 4.2 and 4.3. But other inevitable uncertainties just stem from the general fallibility of human knowledge, and thus affect MSMs and ASMs equally. In this section, I will ignore the former differential problems and assume that MSMs and ASMs face the same degree of uncertainty. The question then is whether MSM and ASM offer different strategies for dealing with such inevitable uncertainty, and which of these strategies is better.

Consider the following example from Eubank et al. (2004). The contact data on which the simulation is based gives a detailed account of social interaction. The model lacks any account of how these social contacts may change under external shocks. The arrival of a threatening epidemic is, arguably, such a shock: it may well have important influence on how often people appear in public, go to work, or go to the hospital. The authors deal with this uncertainty as follows.

One of the most important assumptions in any smallpox model is whether infectious people are mixing normally in the population. [...] We undertook to model two (probably unrealistic) extreme cases: one in which no one who is infectious is mixing with the general population and another in which no one's behavior is affected at all by the disease. In addition, we modeled one more realistic case between these two extremes. (Eubanks et al. 2004, Supplement, p. 11)

The model results strongly depend on the different assumptions. In particular, if people withdraw to the home, then all vaccination policies yield similar results, particularly if there is a delay in the vaccination procedure. However, if people do not withdraw, then LV is substantially less effective than either MV or TV (Eubanks et al. 2004, p. 182, Figure 4).

Note that the MSM here only allows a qualitative distinction: depending on whether withdrawal occurs “early,” “late,” or “never,” the simulation results in a different cumulative number of deaths. Such an analysis is similarly feasible with ASMs. The MSM authors do not assess the uncertainty included in these qualitative results beyond dis-

playing them. While I agree that this seems the correct procedure in this case—as not enough evidence is available to provide a quantified assessment of the behavioral changes under shocks—the question then is why one would go through the additional efforts and costs of creating an MSM, if similar results could have been obtained with an ASM.

What MSMs often aspire to achieve instead is an overall quantification of the uncertainty involved. Although Eubanks et al. do not do this (correctly, I believe), they could have tried to specify a probability distribution over the different behavioral mechanisms and then represent the model outcome as expected cumulative deaths. Such one-size-fits-all approaches in MSMs have been justly criticized for providing *false precision*:

if uncertainty is represented and reported in terms of precise probabilities, while the scientist conducting the analysis believes that uncertainty is actually ‘deeper’ than this — e.g. believes that available information only warrants assigning wide interval probabilities or considering an outcome to be plausible — then the uncertainty report will fail to meet the faithfulness requirement; it will have false precision. (Parker and Risbey 2015, p. 4)

My argument here is that, in most applications of MSMs for policy purposes, non-quantifiable uncertainties arise. These should not be patched over by false precision, as described in the quote above. Alternatively, MSMs are used for providing qualitative results, like the Eubanks et al. example above—which also could have been provided by an ASM. Defenders of MSMs might reply that such qualitative results from MSMs are more accurate than the comparative results from ASMs. However, my earlier arguments in Sects. 4.2, 4.3 and 4.4 question whether this is typically the case. Consequently, the uncertainty quantification strategies facilitated by MSMs are not necessarily better than the strategy of ASMs.

5 Conclusions

In this paper, I argued first that MEMs often suffice for policy purposes, and that second, MSMs are more likely to mislead the model user in making incorrect inferences than ASMs.

In particular, regarding the first claim, MEMs suffice for policy making if they isolate a reliable mechanism that the policymaker can use to systematically influence the system in question. Whether a particular MEM is a valid isolating model requires empirical evidence. But amongst candidate models, those MEMs are most promising, I have argued, which are robust with respect to the idealizations of the isolated mechanism itself. Such robustness investigations, I have further argued, only requires the use of ASMs, not MSMs. MSMs therefore are often not necessary for policy purposes.

Secondly MSMs, although not necessary, might nevertheless appear to be highly convenient for a policymaker. Against this impression, I argued that MSMs rarely exhibit a quality that would actually serve the convenience of the policymaker. In particular, I argued that MSMs might pose more severe problems than ASMs in determining the accuracy of the model; that MSMs might have pose severe problems than

ASMs in dealing with inevitable uncertainty; and that MSMs might pose more severe problems than ASMs with misinterpretation. This of course does not exclude that some MSMs provide good justifications for policy decisions (and even better justifications than some ASMs); but it should caution against a general preference for MSM over ASMs for policy decision purposes.

Acknowledgements Earlier versions of this paper were presented at the University of Helsinki, the University of Bern and the Academy of Sciences and Humanities in Hamburg. I thank the audiences at those events for lively discussions and helpful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alfi, V., Cristelli, M., Pietronero, L., & Zaccaria, A. (2009). Minimal agent based model for financial markets I. *The European Physical Journal B*, 67(3), 385–397.
- Antoulas, A. C., Sorensen, D. C., & Gugercin, S. (2001). A survey of model reduction methods for large-scale systems. *Contemporary mathematics*, 280, 193–220.
- Atkinson, W., Hamborsky, J., McIntyre, L., & Wolfe, S. (Eds.) (2005). Smallpox. In: *Epidemiology and prevention of vaccine-preventable diseases* (The Pink Book) (9th Ed., pp. 281–306). Washington DC: Public Health Foundation.
- Bicknell, W. J. (2002). The case for voluntary smallpox vaccination. *The New England Journal of Medicine*, 346(17), 1323–1325.
- Burke, D. S., Epstein, J. M., Cummings, D. A., Parker, J. I., Cline, K. C., Singa, R. M., et al. (2006). Individual-based computational modeling of smallpox epidemic control strategies. *Academic Emergency Medicine*, 13(11), 1142–1149.
- Cartwright, N. (1994). *Nature's capacities and their measurement*. Oxford: Oxford University Press.
- Cartwright, N. (2007). The vanity of rigour in economics: Theoretical models and Galilean experiments. In N. Cartwright (Ed.), *Hunting causes and using them: Approaches in philosophy and economics* (pp. 217–261). Cambridge: Cambridge University Press.
- Dawid, H., & Fagolio, G. (2008). Editorial. *Journal of Economic Behaviour & Organization*, 67, 351–354.
- Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, 4(5), 41–57.
- Epstein, J. M. (2006). Remarks on the foundations of agent-based generative social science. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics* (pp. 1585–1604). Amsterdam et al.
- Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M., Srinivasan, A., Torocszai, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429, 180–184. See supplement at <http://www.nature.com/nature/journal/v429/n6988/extref/nature02541-s1.htm>. Last Accessed 2 October 2017
- Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256), 685–686.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristics: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1), 107–143.
- Gräbner, C., Bale, C. S., Furtado, B. A., Alvarez-Pereira, B., Gentile, J. E., Henderson, H., et al. (2017). Getting the best of both worlds? Developing complementary equation-based and agent-based models. *Computational Economics*. <https://doi.org/10.1007/s10614-017-9763-8>.
- Grüne-Yanoff, T. (2011a). Isolation is not characteristic of models. *International Studies in the Philosophy of Science*, 25(2), 1–19.
- Grüne-Yanoff, T. (2011b). Agent-based models as policy decision tools: The case of smallpox vaccination. *Simulation and Gaming: An Interdisciplinary Journal*, 42(2), 219–236.
- Grüne-Yanoff, T. (2016a). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, 32(3), 463–483.

- Grüne-Yanoff, T. (2016b). Seven problems for massive simulation models. In M. M. Resch, A. Kaminski, & P. Gehring (Eds.), *The science and art of simulation*. Berlin, New York: Springer.
- Halloran, M. E., & Longini, I. M. (2003). Response. *Science*, 300(5625), 1503–1504.
- Halloran, M. E., Longini, I. M., Nizam, A., & Yang, Y. (2002). Containing bioterrorist smallpox. *Science*, 298(5597), 1428–1432.
- Hine, D. (2010). *The 2009 Influenza Pandemic: An independent review of the UK response to the 2009 influenza pandemic*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/61252/the2009influenzapandemic-review.pdf. Last Accessed 2 May .2016.
- Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 169(3), 615–626.
- Kaplan, E. H. (2004). Preventing second-generation infections in a smallpox bioterror attack. *Epidemiology*, 15(3), 264–270.
- Kaplan, E. H., Craft, D. L., & Wein, L. M. (2002). Emergency response to a smallpox attack: The case for mass vaccination. *Proceedings of the National Academy of Sciences*, 99(16), 10935–10940.
- Kaplan, E. H., Craft, D. L., & Wein, L. M. (2003). Analyzing bioterror response logistics: the case of smallpox. *Mathematical Biosciences*, 185(1), 33–72.
- Kaplan, E. H., & Wein, L. M. (2003). Smallpox bioterror response. *Science*, 300(5625), 1503.
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115(772), 700–721.
- Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science*, 61(3), 541–567.
- Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 41(3), 253–262.
- Lloyd, E. A. (Ed.) (2010). Confirmation and robustness of climate models. *Philosophy of Science*, 77(5), 971–984.
- Mäki, U. (1992). On the method of isolation in economics. In C. Dilworth (Ed.), *Idealization IV: Intelligibility in science* (pp. 319–354). Amsterdam: Rodopi.
- Muldoon, R. (2007). Robust simulations. *Philosophy of Science*, 74(5), 873–883.
- Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44(1), 190–204.
- Parker, W. S., & Risbey, J. S. (2015). False precision, surprise and improved uncertainty assessment. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2055), 20140453.
- Schilders, W. H., Van der Vorst, H. A., & Rommes, J. (2008). *Model order reduction: Theory, research aspects and applications* (Vol. 13). Berlin: Springer.
- Vynnycky, E., & White, R. (2010). *An introduction to infectious disease modelling*. Oxford: Oxford University Press.
- Weisberg, M. (2012). *Simulation and similarity: Using models to understand the world*. Oxford: Oxford University Press.
- Wren-Lewis, S. (2018). Ending the microfoundations hegemony. *Oxford Review of Economic Policy*, 34(1–2), 55–69.
- Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, 44(1), 41–61.