Minding Norms in an Epidemic Does Matter

Klaus G. Troitzsch[1*]

[1*]Institute of Information System Research, retired from Universität Koblenz-Landau, Germany.

Corresponding author. E-mail: klaus.g.troitzsch@bluewin.ch;

Abstract

This paper tries to shed some light on the mutual influence of citizen behaviour and the spread of a virus in an epidemic. While the spread of a virus from infectious to susceptible persons and the outbreak of an infection leading to more or less severe illness and, finally, to recovery and immunity or death has been modelled with different kinds of models in the past, the influence of certain behaviours to keep the epidemic low and to follow recommendations of others to apply these behaviours has rarely been modelled. The model introduced here uses a theory of the effect of norm invocations among persons to find out the effect of spreading norms interacts with the progress of an epidemic. Results show that norm invocations matter. The model replicates the histories of the COVID-19 epidemic in various region, including "second waves", and shows that the calculation of the reproduction numbers from current reported infections usually overestimates the "real" but in practice unobservable reproduction number. Note: This version is preliminary. Before a more consolidated version can be written, more information about the autumn 2021 wave and potential later waves of the pandemic at least in some parts of the world is necessary.

Keywords: norm invocation, salience of norms, infectious desease, epidemic, pandemic

JEL Classification: I12

MSC Classification: 60 , 68U20 , 65C05 , 65C20

1 Introduction

This paper tries to shed some light on the mutual influence of citizen behaviour and the spread of a virus in an epidemic. While the spread of a virus from infectious to susceptible persons and the outbreak of an infection leading to more or less severe illness and, finally, to recovery and immunity or death has been modelled with different kinds of models in the past, the influence of certain behaviours to keep the epidemic low and to follow recommendations of others to apply these behaviours has rarely been modelled.

Although in the current Covid-19 pandemic nowhere near all of the parameters determining the spread of Covid-19 are known or, at least, estimated, it seems that standard models of classical epidemiologiy are not applicable to this global event. The spread of SARS-CoV-1 in 2003 could still be sufficiently well model with the standard SIR model leading to a good approximation of empirical data to the logistic curve (see Hsieh, Chang, & Lee, 2004; Zhou & Yan, 2003) although Hsieh et al. found that truncated data underestimated the total number of infected persons but still was a well estimated with a logistic curve. This soon turned out to be different in the case of Covid-19 where first attempts to estimate a logistic curve before the turning point was reached seemed promising (although estimating the date of the inflection of the curve yielded later days for every new estimation with additional data — as was the case for Hsieh et al.): After the inflection point, when the rate of increase of cumulative case numbers had reached its maximum and the disease started to decline in a country the decrease turned to be much slower than the increase had been, leading to the typical assymmetric form of the graph of daily newly-infected persons in a country. This observation was the reason why the model presented in this paper was first designed to find out which non-biological causes could have had the observed effect. It seems that one cause is the fact that human populations are much less homogeneous than the SIR and related models on the macro level and even some models on the micro level assume. Moreover, humans can take precautions to be spared from an epidemic and ask others to take such precautions, too, let alone the competence of state institutions to impose a lockdown on part of the population.

The next section tries to give an overview of the models proposed during the covid-19 pandemic during spring and summer of 2020, an overview which will necessarily incomplete as only few have been made available over the Internet, and only very few had been properly published at the time this paper was written. Section 3 describes the current state of the norm-oriented discrete-event simulation model of a pandemic similar to covid-19, as a model of this pandemic cannot yet be designed as too many parameters are still only vaguely known. Section 4 gives an overview of the results of a number of simulation runs with varying parameters, whereas Section **??** discusses the model and its results, comparing them to other recent models.

The research questions which this model attempts to answer are the following:

norm orientation: Does it make a difference when people in a community are more or less aware of others' behaviour with respect to risk avoidance and shape their own behaviour according to what they observe in their neighbourhood? And does this norm orientation change over time, and if so, how will it change under the risk of an epidemic?

settlement structure: Does settlement structure make a difference, i.e. do different allocations of agents to subregions of the simulated world lead to different pandemic histories?

heterogeneity of risk aversion: Does it make a difference when people have different degrees of conviviality and carelessness which make the avoid or seek risky situations?

person-zero assumption: Can current pandemic histories be explained with the assumption that one infected person intrudes a community?

mutation: Can current pandemic histories be explained with the assumptions that mutations do not affect infectiosity?

reliability of estimates of R_0: Is the estimate of the basic reproduction number R_0 with the nowcasting method an der Heiden and Hamouda (2020) reliable, and if so, under which conditions?

reliability of estimates of the percentage of positive tests: Is the percentage of positive tests on all tests (the test positivity rate, see Wu et al. (2020)) a reliable measure of the severity of a pandemic?

vaccination: What is the effect of a late start of vaccinations and of the availability of doses, i.e. how long it takes until the total population can have had an opportunity to be vaccinated?

The first of these questions is the central topic of this paper, answers to the other questions are mere byproducts of a model which tries to replicate the course of the pandemic in small regions of German speaking Central Europe. That norms seem to make a difference was empirically fairly well conformed in Gelfand et al. (2021), a study comparing the normative culture in nearly all countries all over the world with their success or failure to cope with SARS Covid-19 — this paper aims at deepening our understanding of how norms work on a micro level.

2 Review of Current Agent-Based Epidemic Modelling

Whereas modelling in epideniology goes back to the 1920s, although based on systems of differential equations — models which at best are valid for large homogeneous population in which everybody can infect everybody else[1] —, individual-based approaches to modelling epidemics and pandemics are not much older than (Epstein & Axtell, 1996, pp. 138–152). With the growth

[1] One promising approach along these lines — with parameters varied over time according to empirical evidence — was recently published by Agrawal, Kanitkar, and Vidyasagar (2021). Its retrospective and predictive quality turns out to be quite good for most but not all states and union territories of the Republic of India. Detailed results can be found at https://www.sutra-india.in/. For another similar approach, see (Nistal et al., 2021)

of the agent-based modelling community and particularly with the advent of the Covid-19 pandemic, a number of new agent-based models continuing the tradition of Epstein and Axtell (1996) has emerged some of which will be discussed in this section although most of them are still work in progress.

This is also true for the perhaps most recent modelling project called COMOKIT (Gaudou et al., 2020) which combines a model of the transmission of a Covid-19 infection among people moving around on the regional scale of a closed commune and on an hourly time scale; it has already been used for three case studies covering two small towns in Vietnam and one in Southern France (with between 7,700 and 14,000 inhabitants). Agents move between different places and and perform different activities there. COMOKIT contains an intervention model as besides the individual agents there is an authority agent which is informed about the health state of the population and can applies policy to fight the epidemic (for instance "the policy that forbids any activity to any Individual agent who has been tested positive" (Gaudou et al., 2020, p. 18 of the ODD)). The model is "generic, scalable and thus portable in a variety of social and geographical contexts", it is "designed to be modular and flexible enough to allow modellers and users to represent different strategies and study their impacts in multiple social, epidemiological or economic scenarios." (Gaudou et al., 2020, Abstract) and can in priniciple be applied to other communes or subareas and, given sufficient computation power, to even larger geographical areas.

Hoertel et al. (2020) presented an agent-based model discussing the effect lockdowns and similar measures in France in mid-2020. Their model calibrated and validated quite nicely for the first five months of the pandemic in France and tried to predict the result of different lockdown measures (including social distancing and mask-wearing), and used it in (Hoertel, Blachier, Sánchez-Rico, Limosin, & Leleu, 2021) for predictions until February 2021 based on empirical data until September 2020 under different assumptions about the percentage of mask wearers, concluding that "that early implementation of face mask use with an achievable adherence of 80 ¸% in the population would have reduced cumulative COVID-19 incidence, mortality, and hospital-bed occupancy, as compared to the observed epidemiological situation in France, supporting the importance of this protection measure."

The problem of simulating very large artificial societes is dealt with in an older paper Perumalla and Seal (2012) describing an epidemic simulation model appropriate "scenarios representing large population sizes, with mobility and detailed state evolution modeled at the level of each individual, exceeding several hundreds of millions of individuals in the largest cases, are successfully exercised to verify model scalability." The discrete-event oriented reaction-diffusion model[2] of the authors uses machines with up to "65,536 cores of a large Cray XTS system" (Perumalla & Seal, 2012, p. 768) which are usually not available to most researchers — such that the following review is restricted to

[2]Models of this kind are also applied to "to simulate the correct time evolution of ... complex biological system[s]", Figge (2005).

models of small groups of researchers with access only to standard computing facilities.

An approach to replicate the progress of the Covid-19 in Northern Italy was programmed by Pietro Terna and his colleagues (Pescarmona et al., 2021; Terna et al., 2020). This model uses the topography of Piedmont with its 4.35 million inhabitants reduced to 4,350 person agents: "the model is related to the Piedmont scale, with 4,350 agents vs. 4.35 millions of inhabitants. The scale 1 to 1000 is over-represented in the case of schools, with their classrooms with a realistic number of students, apartments with a realistic quantity of inhabitants, and likewise workspaces, hospitals, nursing homes." It allows for various interactions between model and user, and it also tries to model the places where infections can occur in a nutshell. The model description in the NetLogo file discusses "considerations on dimensions" which are necessary as the model offers a drastic downscaling of both population and area (25,400 $km^2 \rightarrow 2601$ patches subdivided into nine smaller squares, hence one km^2 is represented by approximately one of the 23,409 smaller squares within the patches).

Finally, an approach published by Vermeulen, Pyka, Müller, and Kugelmeier (2020) deserves to be mentioned which simulates a village with a number or houses, schools, offices and other enterpreises, recreation areas, supermarkets and even a cemetery among which a small number (428 agents) move, meet and potentially infect each other. This model, programmed as a "policy laboratory for COVID-19 containment strategies", makes easily visible what happens in such a small community and allows for two dozen strategies in six different areas to show users (obviously not only politicians and administrators but also grammar school pupils) what might happen if, for instance, gatherings were limited, incoming travellers were not allowed, hospitals would only admit critical cases, all people worked from home if possible and/or schools were closed when the first infected puoil was detected, and how the epidemic would develop under different assumprions of infection risk and immunity period if the measures taken were or were not accepted and/or working from home appeals were or were not followed.

Generally speaking, most of the models of pandemics and particularly of Covid-19 seem to be written in order to understand what could happen during the progress of a pandemic; some, in fact, also try to forecast and/or to advise policy making. The discussion whether predictions or forecast of this pandemic is possible and whether models of this kind are necessary and/or useful continues. It was initiated by a "call to action" (Squazzoni et al., 2020) and continued, among others, by (Steinmann, Wang, van Voorn, & Kwakkel, 2020). This paper does not aim at predicting at all; its main purpose is to analyse whether replications are possible and why they might fail.

3 The Model

This is an event-oriented model of an epidemic in which agents move around, eventually infect each other, fall ill, recover, become immune or die from the disease. Agents sense how many infectious other agents are in their neighbourhood and all over the population and calculate their personal risk of being infected depending on their own (constant) conviviality and carelessness and on their (variable) risk perception.

3.1 The artificial world

3.1.1 Area and population

The model covers an area of 10,201 (101 × 101) patches, divided in several (up to 13) regions (`subareas`) with equal population and different population density (the smallest area has the highest density). Depending on the different population density and the numbers of agents , the distribution of the minimal distance between nearest neighbours (measured in patch length) can be taken from Table 1. The simulated world of the model does not try to replicate a certain region (as is the case in Terna et al. (2020) or Gaudou et al. (2020)): with different seeds of the random number generator, agents are placed in different arbitrary manners over their world and its `subareas`.

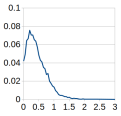
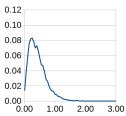
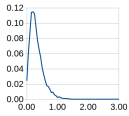
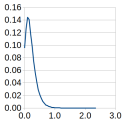
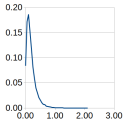
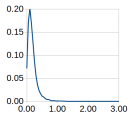
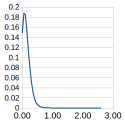
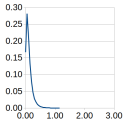
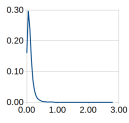
The mechanism of spreading the agents over the simulated world first assigns an equal number of agents to each `subarea` and then selects a number of patches within each `subarea` and spreads agents around these patches. such that in the end the model world is populated with 10,000 or 38,000 or 70,000 agents.[3] The topographical distribution is in this case sufficiently realistic, as the diagrams in Table 1 show.

Table 1 — with respect to the settlement structure — shows that the variants of the model cover scenarios which correspond to villages and small towns in subareas[4] or small countries or territories such as Gibraltar (33,689), San Marino (33,945), Liechtenstein (38,149) (these three being in the range of the model runs with the medium total population) or Andorra (77,289) which can be taken to calibrate the model with respect to the duration and impact of the empirical epidemic in these villages, towns, subareas or small countries or territories.

[3] ′To calibrate the distribution of the minimum distance between an agent and its nearest neighbour, a region in Northern Germany — a square with its center at 51°54′ N, 9°32′E, an edge length of about 38 km and about 72,800 inhabitants between the towns of Bodenwerder, Holzminden and Stadtoldendorf — was used (the population numbers for the places on this map were extracted from Wikipedia in December 2020). The map used is available from https://numis.niedersachsen.de/kartendienste. In this stylised-fact variant one patch corresponds to a square of approximate $375 \times 375\,\mathrm{m}^2$, and the number of persons living in such a square is assumed to be proportional to the number of buidings identifiable on a 1:25,000 map within this square.

[4] 15 out of the 294 German *Landkreise*, seven out of 26 Swiss cantons and 55 out of 94 Austrian *Bezirke* have less than 74,245 inhabitants, 15/4/15 even less than 35,000; and in France, too, 81 out of 332 *arrondissements* have numbers of inhabitants that are in the range of the simulations of this model. Hence even the model runs with the medium total population have the size of many of these territorial subdivisions.

Table 1 Mutual distance of agents in the simulation and in stylised-fact real regions

total population	model	region around Holzminden[a]	region around Basel[b]
10,000			
38,000			
70,000			

[a] See footnote 3

[b] The region around Basel is centered on a point called Steinenschanze (about 500 m north of the main station) and is a square of about 18 km edge length.

3.1.2 Output of the model

Plots and monitors keep track of the epidemic. Some examples are given in Section 4. In general, the progress of the pandemic can be followed in plots showing the cumulated numbers of agents susceptible, infected, ill, recovered and dead. Other plots show the current distribution of instance variables of agents, such as the salience of norms, the product of conviviality and carelessness of infected and susceptive agents, the perceived risk and the number of successful infections from all currently infectious agents. Another plot shows the progress of parameters of the pandemic as the basic reproduction number (see Subsection 4.5.2) and the overdispersion parameter (Subsection 4.5.3). A detailed description of the model along the lines of the ODD protocol (Grimm et al., 2006, 2010; Müller et al., 2013) can be found in the Appendix C and/or at DOI 10.25937/7vkh-tt08.

3.2 Agents

3.2.1 Agents' features

At the very beginning all agents are scheduled to move a short distance within some short time. All (except one `zero-person`) start in the `state` "susceptible", and their infectiosity attribute `infectious?` is `false`. If at the end of this move they do not encounter an infectiuos agent they move on after some time, otherwise they take the risk of getting infected. When an agent is infected, its `state` changes to "infected". It continues to move around, become infectious themselves — its infectiosity attribute becomes `true` with a small delay. Some agents remain asymptomatic, others change their `state` to "ill"

and are moved to the center of their `subarea` — as it were, a hospital. After they recovered or died, they return to their start patch. If they recovered they continue to move, lose their infectiosity after some time, become immune for a limited (but rather long) period, and if they lose their immunity they become susceptive again.

Infections are reported with some delay, but at the very latest when an agent falls ill and is brought to hospital.

The individual risk of being infected during an encounter depends, of course, on how many infectious agents are in the neighbourhood (defined by `distance-for-infection`, but also on three attributes of the individual agents:

- `conviviality`, a beta distributed random number between 0 and 1 initialised in the setup and kept constant
- `carelessness`, another beta distributed random number between 0 and 1 initialised in the setup and kept constant
- `perceived-risk` which at any time is the smaller of the percentage of infectious agents within a radius of 5 and the overall percentage of infectious agents

The risk of being infected is proportional to the product of the current infectiosity of the virus in the infector (`virus-infectiosity`, which is initialised as `infections-per-contact` and modified according to `mutation-rate` whenever an infection is transmitted), and the individual attributes `conviviality`, `carelessness` and `perceived-risk`.

3.2.2 Events in which agents act or are acted upon

In the standard version of the model (the normative variant, the normative mode can be switched off) the following events can be scheduled by the model agents:

- `move-around`: the agent moves to another place
- `set-infectious`: the agent is infected and becomes infectious
- `report-a-positive-test`: the agents is reported to the observer as infected
- `fall-ill`: the agent shows symptoms and is brought to hospital
- `recover-or-die`: the agent either recovers or dies and is returned to its original patch
- `report-a-death`: the death of the agent is reported to the observer
- `become-immune`: the agent is no longer infectious nor susceptible
- `lose-immunity`: the agent loses its immunity and is susceptible again

For more details see Figure C1 in the ODD in the appendix.

Three switches allow the introduction of additional effects:

additional infections: `infections from abroad` allows for introducing additional infections (in a real-world scenario these would be caused by people who return from a region abroad which could be classified as a particularly risky region).

superspreading events; `superspread?` can switch on or off whether some agents — those with high values of `conviviality * carelessness` — organise events by asking a group of neighbours to convene in a very small region around their current places where they are potentially exposed to a high risk of being infected — if infectious agents happen to be around.

lockdown: `auto-lockdown` when the infection rate exceeds a standard value (for instance 50 new infections per 100,000 during the past seven days). This leads to a sudden decrease of the mobility of the agents until the infection rate falls below another standard value.

3.3 Baseline scenario

When the model starts, exactly one agent in the subarea with the smallest density (alternatively: with the highest density, or at a random patch) is infected, and the `mutation-rate` is 0. This scenario is not actually realistic, as in real territories people from abroad, some of them infected, and not only one `person-zero`, will enter, or people will move out and return (some of them infected). But for a first study of alternatives to classical equation-based models, this baseline scenario might be a useful illustration of the difference between a macro model and its stochastic individual-based alternative.

The initialisation for the run shown in Fig. 1 (and also for all other runs shown in later figures as discussed in subsections 3.4.1 and 3.4.3) can be found in Tables B2 and B3.

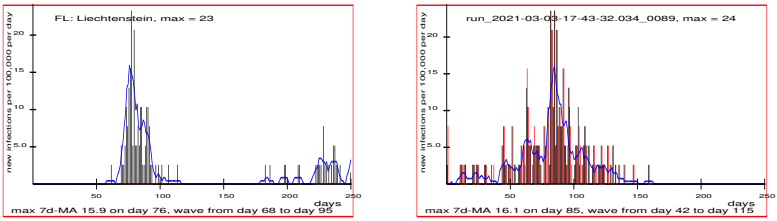


Fig. 1 Comparison of the pandemic in the Principality of Liechtenstein (first 250 days of 2020) and a baseline run in a population of the same size (38,149)

The pandemic history of the baseline scenario — see Fig. 1 — usually shows only one wave, and if there is a second wave, its peak is much less severe than the first wave's. This is also true for a majority of the runs used for sensitivity analysis, calibration and validation — these runs were excluded form part of the comparisons. This is, of course, in line with predictions from equation-based models which can have only one wave in the simple case. The main difference from these is the fact that the rise of the baseline-scenario pandemic is usually steeper than the decline, and this could also be observed in the first waves of some territories — see the example in Figure 1.

The difference between the real data of Liechtenstein and the exemplary baseline run is mainly that the simulated pandemic started earlier (just because the zero person was infected on the first day of the year whereas the Liechtenstein zero person entered the region only seven weeks later) and that the simulated pandemic lasted longer although on a very low level whereas Liechtenstein seems to have been invaded by another zero person after more than two months (23/04 till 01/07) without any new infections. The baseline version of the model would then stop as it does not foresee any infection from outside the model region.

It should be noted here that out of 100 runs with the same parameter combination, only eight produced more than 100 infections over the whole run and an incidence rate peak over 10, with these two parameters correlated with an R^2 of 0.132.

Figure 1 also shows that in real cities and subareas in Germany, Austris, Switzerland and the Principality of Liechtenstein (with the latter as an example) the duration of the first wave was much shorter than in the baseline scenario — an observation that led to several more realistic scenarios in which additional infections occur during part of the simulation time. Moore, Lipsitch, Barry, and Osterholm (2020) showed three likely scenarios for the COVID-19 pandemic derived from the 1918-1920 influenza pandemic which they called "peaks and valleys" with peaks and valleys with equal size and a period of about five months, "fall peak" with a high peak forecast for October 2020 and smaller peaks in April/May 2020 and again in the autumn of 2021 (similar to what can be observed at the time of writing this) and "slow burn" much like the first scenario, but with the exception of a higher peak in April/May 2020 and smaller peaks for the rest of the 28 months forecast period.

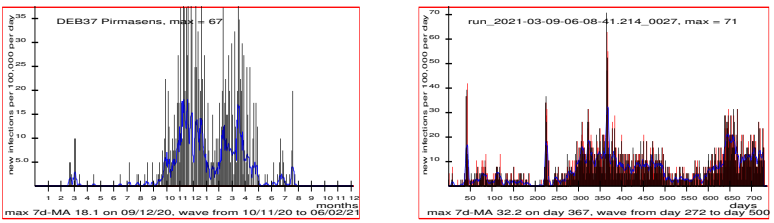


Fig. 2 Comparison of the pandemic in Pirmasens (all available data, 40,231 inhabitants) and a halfway similar run in a population of 38,149 — starting point for the sensitivity analysis

The simulation model should at least be able to replicate these three scenarios — as Fig. 2 shows as here the similarity between the empirical course of the pandemic (this time with a town in Rhineland-Palatinate and some surrounding villages) and a simulated course is much better than with the baseline — but here, too, the similarity is still poor, and the selection of the empirical history and the simulation run might seem arbitrary. All German,

Austrian and Swiss regions have in common that a period of three to four months between the first and second waves shows only very few new infections, whereas the simulation runs either have only one wave (and then the epidemic is over) or the period between the first and second waves still shows a moderate incidence — which is also true for the period between the second and third waves in the German and Swiss regions. Nevertheless, parameter combinations used in the run depicted in Fig. 2 will serve as starting points for calibration and sensitivity analysis.

3.4 Special scenarios

3.4.1 Scenarios with new infections from outside during the epidemic

In this scenario, additional infections are scheduled for some agents and for some periods during the simulation run. Following the assumption that outside a certain territory other territorities are also subject to the pandemic, that in the real world people move between territories and that this happens more frequently when people go abroad for their holidays (which in Europe is typically February and summer, hence additional infections are scheduled for the months of February, June, July, and August in some of these scenarios. The model provides the following options with a fixed rate (`travellers`) of people returning from abroad:

holidays: returns from abroad happen only in February and July to September,

continuous: returns happen at any time of the year,

seasonal: a certain percentage (`%continuous`) of the returns happen at any time of the year, the other half happens in February and from July to September,

others: like the holidays option, but with different collections of months.

3.4.2 Lockdown and quarantine

In this scenario, measures are taken by the observer (either programmed or by intervention of the modeller) to reduce the infection risk for all agents, both by reducing the radius of their movement and by asking them to wear masks or stay at home. The latter measure uses the normative memories of the agents by the respective norm invocation which in this case does not come from neighbours but from the program or the modeller. The automatic lockdown starts when the seven-days incidence rate exceeds the value selected by the respective slider and ends when this rate falls below half of this value. The threshold for the automatic quarantine (no further movement is allowed)[5] is four times the lockdown threshold.

[5] Quarantine was never used in the simulation runs reported here.

3.4.3 A scenario with superspreading events

In this scenario, some agents are scheduled to invite other agents in their neighbourhood to come together for a couple of hours. During this time they are exposed to transmission to a greater extent than in ordinary times. Such superspreading events can, of course, can also occur during lockdown and quarantine. To be more precise, the first such event is scheduled for some time during the first two weeks, and it is organised by one of the agents with the highest value of the product of `conviviality` and `carelessness`. Further events follow at some time during the following four weeks, and they are organised by the same or another agent with the highest value of the product of these two variables. When the event starts it convenes, the up to 50 agents within a radius of 10 patches and with the highest values of the product of `conviviality` and `carelessness` move to a region with a radius of half a patch where they are exposed to transmission from infectious agents if there are any. The event then lasts between three and eight hours, and afterwards all agents go back to the point where they were initialised (their `home`).[6]

3.4.4 A scenario with virus mutation

In this scenario, the virus starts with an fixed initial infection rate (the probability to be infected as a consequence of a contact with an infectious person). With every new transmission to another susceptible agent, the infection rate ρ_t of the virus decreases or increases with a random number $\mu_t \approx U(-\mu, \mu)$. At the same time the probability to remain free from symptoms α increases or decreases with the same rate such that

$$\rho_{t+1} = \rho_t + \mu_t \tag{1}$$

$$\alpha_{t+1} = \alpha_t - \mu_t \tag{2}$$

This is a simplification of the current assumption that new mutations are more contagious but may lead to less or milder symptoms (see for instance Lauring and Hodcroft (2021); Young et al. (2020)). The case of the delta mutant — which is both more infectious and can lead to more dangerous symptoms can also be modelled with a slight change of the code.

3.5 The influence of norms on the behaviour of the agents and on the progress of the pandemic

The influence of norms on the behaviour of agents can be analysed with the help of a mechanism developed in Conte, Andrighetto, and Campennì (2013) (for details see Appendix A) which controls the propensity to take different actions in a twofold manner: this propensity depends on two "drives", one of which takes into account utility considerations — in this model the perceived risk to be infected — and normative considerations — in this model

[6] This scenario is not discussed in the current version of the paper.

the norm invocations received from others. The actions which depend on these two considerations are the following:

- wearing masks (as opposed to breathe freely) which reduces their probability of being infected by the factor 0.2 (can be changed in the code) and
- staying at home (instead of going to places where they can be infected) which reduces their probability of being infected to 0.

Switching between either behaviour corroborates slightly the propensity to keep on with the new behaviour; neighbours are encouraged to behave in the same manner.

All agents are initialised with a memory of past norm invocations (for the theory behind norm invocations and norm-dependent behaviour see Conte et al. (2013) and Elsenbroich, Anzola, and Gilbert (2016)). The number of such initial norm invocations depends on the global variable `BACKGROUND` (-10 .. +10) which determines the mean of the number of invocations of norms relating to a fear of the pandemic and the fear of losing one's freedom (the individual numbers are beta distributed around this mean with a range of 20). The norms are the following:

- `FEAR OF PANDEMIC` related:
  - `NORM: STAY AT HOME`
  - `NORM: WEAR A MASK`
- `FEAR OF OPPRESSION` related:
  - `NORM: GO TO PARTY`
  - `NORM: PUT OFF MASK`

Agents calculate the salience of these norm and the propensity to take repsective actions:

- `PUTTING ON MASK`
- `PUTTING OFF MASK`
- `STAYING AT HOME`
- `GOING TO PARTY`

The salience of a norm depends mainly on how often an agent has already abided by the norm and how often it violated this norm, but also how often it has seen other agents abiding by the norm or violate it.

The respective propensity depends on both the *individual drive* and the *normative drive*. The former depends only on the perceived risk compared to the perceived risks of the agent's neighbours, whereas the latter depends on the current salience of the four norms (for more details see A). Finally, it is the normative drive weight `NDW` which decides whether the pure risk calculation or the invocations from the neighbourhood prevail. Hence it is the combination of the input parameters `BACKGROUND` and `NDW` which allows a theory-grounded statement whether normative behaviour matters in a pandemic.

4 Results

4.1 Input parameters and output metrics analysed

The following analysis restricts itself to the following input parameters:

- `mutation-rate`, the increase of the infectiosity of the virus when it is transferred from one agent to another: either 0.0 or 0.05; this leads to an overall increase from the start value of 0.05 up to values (averaged over all currently infectious agents) slightly below 1.0,
- `mean-distance-to-nearest-neighbour`, a parameter derived from the topographical distribution of the agents over the whole area — the main difference is between the runs with 38,000 and 70,000 agents,
- `BACKGROUND`, in case of norms playing a role, a predisposition common to all agents determining which risk they prefer to avoid: the risk of loosing their freedom (`BACKGROUND` < 0) or the risk to get infected (`BACKGROUND` > 0),
- `NDW`, the weight of the normative drive as opposed to the individual drive to to take any precautions against being infected (`NDW` $= 0$ means that the behaviour of the agents depends only on the perceived risk of being infected, not on any norn invocation),
- `lockdown-at`, a global parameter — an incidence rate — which when `auto-lockdown` is set to either `internal` or `external` determines that the acitivity radius of all agents is decreased when this incidence rate is exceeded.

Additionally two derived parameters are used:

- `mean-daily-moves`, a parameter derived from the combination of the individual attributes `carelessness` and `conviviality` which depend on the distribution of these two attributes and the global parameters
  - `long-distance`, `short-distance` and `far-commuters`, the latter being the rate of agents whose maximum travelling distance is `long-distance` whereas the others activity radius is `short-distance`,
  - `infections-from-abroad`, the number of infections caused by agents which (virtually) returned from abroad (their percentage being `travellers`, a constant of 0.015), but whether they are infectious and whether there infect neighbours at the time of their (virtual) return depends on their own history and the situation at the place to which they return (if at the scheduled time of their return they have recovered from an earlier infection they can no longer infect, and their neighbours' risk of being infected is zero.

These depend partly on the history of each simulation and are, in a way, path-dependent and slightly correlated with the input parameters proper.

A number of other input parameters mainly describing the features of the epidemic are listed with their standard values in the ODD+D description in Appendix C.

All the other input parameters are neglected here, as this paper concentrates on the questions detailed at the end of the Introduction and thus on the potential influence on heterogeneity of risk avoiding, risk seeking and normative behaviour on the progress of an arbitrary epidemic.

The output metrics analysed are the following (numbers of agents, infections and death are per 100,000):

- the maximum number of daily new infections over the whole simulated period,
- the maximum number of deaths per day over the whole simulation period,
- the total number of people agents which were ever infected,
- the total number of deaths,
- the number of days until these two maxima are reached — both in the simulation runs and the empirical time series, several waves of the epidemic can be observed. In rare empirical cases and a number of simulation runs it was the first wave which had the highest peak (see also Figure 4), the calculation of the means hides part of the real distribution of this output metric.

As these five variables are highly correlated, only the results for the last three are reported in Table 2.

All five output metrics are also available for all NUTS-3 regions of Austria, Germany, Switzerland and for several small territories and coutries elsewhere, to allow a comparison between simulations runs and empirical epidemic progress (where the empirical data could only be used until this paper was finished, hence only taking into account about 20 months).

4.2 Sensitivity analysis

The first observation from simulation runs which are not reported here is: All input parameters not mentioned above had no significant effects on the output parameters — at least not for value combinations that led to simulation results which were similar to empirical time series of the German speaking countries analysed.

Table 2 shows the effects of the input parameters on the output metrics. The global effects are measured in terms of multiple regression coefficients and the effects of individual input parameters are given in terms of standardised regression coefficients (which are nearly the same as correlation coefficients as the correlations between most input parameters are zero).

The first conclusion that can be drawn from Table 2 refers to the columns a and b of this table, as it shows that lockdowns make a difference to all three analysed output metrics. Early peaks do not seem to lead to any lockdowns (it seems that the causality is the other way round here); a high rate of infections from abroad and a long distance between nearest neighbours seem to delay the peak in the absence of lockdowns. Lockdowns reduce the number of infections and of deaths by about 11 per cent, whether the lockdown is started at a low or high incidence rate makes only a small difference.

Table 2 Effects of hetereogeneity mode and normative predisposition on the output metrics — all 900 simulation runs

input parameter	range	standardised regression coefficients					
		peak day[a]		deaths[b]		cases[c]	
		a	b	a	b	a	b[d]
mutation rate	{0.00, 0.05}	n.s.	0.118	0.571	0.592	0.592	0.642
mean distance to nearest neighbour	[0.1682, 0.2325]	0.611	0.567	-0.048	-0.104	n.s.	-0.041
mean daily moves	[0.5592, 0.6166]	n.s.	0.106	0.285	0.229	0.277	0.259
infections from abroad	[128.6, 1448.6]	0.218	0.358	0.490	0.498	0.473	0.457
normative background	{−6, 0, 6}	n.s.	n.s.	-0.058	-0.034	n.s.	-0.034
normative drive weight	{0.3, 0.6, 0.9}	n.s.	n.s.	-0.190	-0.212	-0.226	-0.245
lockdown start rate	{50, 100}	N/A	n.s.	N/A	n.s.	N/A	0.043
	mean	208.3	240.2	92.4	82.3	3468.7	3061.1
	st dev	173.8	199.9	74.2	62.8	3001.3	2451.8
	R^2	0.458	0.469	0.857	0.834	0.863	0.851

[a]Number of the day when the maximum incidence rate occurred.
[b]Total deaths per 100,000.
[c]Total cases per 100,000.
[d]a: no lockdown, b: lockdown ordered at various incidence rates

4.3 Calibration

The aim of calibrating this model is to find input parameter combinations that resemble best the progress of the Covid-19 (or any other) pandemic in a region of comparable population size. For this calibration, German *Landkreise*, Austrian *Bezirke* and Swiss cantons (subareas, NUTS-3 regions)[7] were used with their data about reported positive tests per day. Some of those with approximately 38,000 or 70,000 inhabitants are presented in Figure 3[8]. These diagrams show that their typical shape shows two or sometimes three maxima of different height referring to the peak of the first and second or third waves. The maximum incidence rate (positive tests per 100,000 within a week) is between 4.77 and 540.34 (first quartile 18.82,.median 25.32, last quartile 37.82, last decile 59.7, last 20-tile 76 such that the distribution is extremely skewed), but as at the time of writing this, the fourth wave in Germany, Austria and Switzerland is not yet over[9]). The peak usually falls in the middle of the overall

[7]The German data used for this calibration were retrieved from the Robert Koch Institute (RKI) (Robert-Koch-Institut, 2021) and the German Statistical Office (Statistisches Bundesamt, 2020). The former is a complete list of all individual cases reported to the RKI. For the comparison these data were aggregated to subarea and day such that this aggregation yielded the history of Covid-19 infections for each of the German NUTS-3 regions (see also Figure **??**). The latter is the annual report of the area, population and population density of the German NUTS-3 regions. The Swiss data were retrieved from (Bundesamt für Gesundheit, 2021), and the Austrian data were retrieved from (Bundesministerium für Soziales, 2021). Both sources keep — among others — large files documenting all tests and deaths reported per NUTS-3 region, age-class and sex. From these, files were aggregated documenting the number of positive tests per NUTS-3 region and day.
[8]The Swiss canton Uri has about 37,000 inhabitants, and the Principality of Liechtenstein has approximately 39,000 inhabitants. In Switzerland this is the canton of Jura with about 73,600 inhabitants, in Germany these are Lichtenfels (66,766), Sömmerda (69,427), Kusel (70219), Holzminden (70,459), and Tirschenreuth (72,046) — the latter had the highest incedence rate all over Germany in the first wave and also high values in the second wave)
[9]This is the reason why this version cannot be complete as at the time of writing another wave is still in progress.

duration of the respective wave, but the distribution of the position of the peak within the highest wave is very flat, has several modes and the second and third quartiles extend from 40 to 64 (when 100 is the time between the first and the last day when the moving average is higher than 10 per cent of the maximum of the wave). The distribution of the peak date can easily be seen in Figure 4.

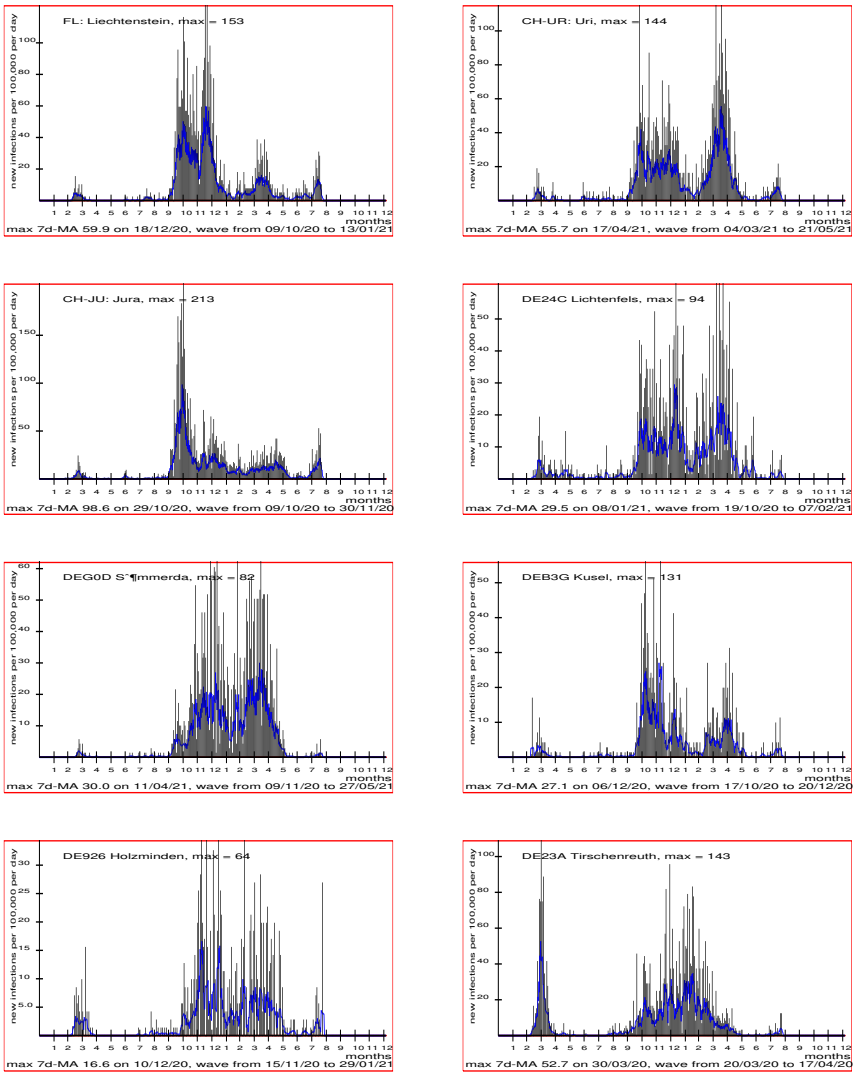


Fig. 3 Progress of the pandemic in the Principality of Liechtenstein, five German and two Swiss NUTS-3 regions with approximately 38,000 and 70,000 inhabitants (excluding NUTS-3 regions which consist of just one town, new infections — positive tests — per day. blue: seven days moving average)

Although there are differences between the eight regions mentioned in Figure 3 it seems that an input parameter combination which can be used as a starting point for sensitivity analysis should lead to

- at least two waves about nine to ten months apart,
- the second wave higher than the first,
- the duration of the second wave between 16 and 19 weeks,
- the peak of the second wave early in its third month, and
- the third wave following the second at a comparably short interval.

This means that among all simulation runs, the input paramater combination must be optimised against these four imprecisely given criteria.

But even in September 2021 it is still too early, and perhaps even impossible, to calibrate the simulation model against counties, cantons and even nations, as the course of the pandemic is multifaceted on both the NUTS-3 and the nation state levels. In Germany, Austria and Switzerland, for instance, one can identify at least five clusters of NUTS3-regions along the time when the highest peak of the epidemic was reached (April 2020, November 2020, Januar 2021, and April 2021 — not taking into account the autumn-2021 wave whose fate cannot be analysed yet) and according to the incidence rate at that time (with a clear difference between Germany on one hand and Austria and Switzerland on the other hand). Figure 4 shows this clustering, and Figure 3 shows examples with the highest peak during the first wave (Tirschenreuth), at different times of the second wave (early: Jura 28/10/2020 and Kusel 07/11/2020, around Christmas: Fürstentum Liechtenstein 19/12/2020, in January: Lichtenfels 08/01/2021, Holzminden 16/01/2021) and during the third wave (Sömmerda, 24/03/2021, Uri 04/04/2021).

From Figure 4 simulation runs can be selected which fit best with German NUTS3-regions. Figures 5 and 6 show a few examples of some of these regions and runs similar enough[10], at least with respect to the height and time of the highest peak. These three figures also display some information about both the simulation runs and the German regions with respect to the two features of the epidemic progress:

- all runs showed several waves, the highest peaks were more or less equally distributed between the first, second and third waves,
- the simulation runs had the peak of their second wave a little earlier than the empirical regions.

Figure 7 shows a comparison between empirical and simulation results, restricted to those simulation runs whose output metrics fell in the range of the empirical date from German NUTS-3 regions and smaller countries and territories all over the world. The small black dots represent regions with a population comparable to the synthetic populations in the simulation runs, the small black circles represent all other regions in Germany (and show at the same time that the difference in the two variables between smaller and more

[10] A dissimilarity matrix, as it was used to select the regions and runs in Figures 5 and 6, is of course more precise.

Table 3 Cluster and country means of the maximum incidence and the peak date in the NUTS3 regions of Austria, Switzerland, Germany and Liechtenstein, compared to 900 simulation runs

		Spring 2020	Autumn 2020 low	Autumn 2020 high	Autumn 2020 extreme	Spring 2021	Total
Austria	incidence mean	8587	6877.9	8570		5132.5	7633.6
	peak date mean	2020/03/21	2020/11/21	2020/11/16		2021/03/01	2020/11/17
	N	1	50	42		1	42
Switzerland	incidence mean		7303	10626.4	77847.7	9176.4	19378.5
	peak date mean		2020/11/18	2020/10/30	2020/11/17	2021/04/17	2020/11/17
	N		12	9	4	1	26
Germany	incidence mean		4411.8	7297.3		4332.5	4639.2
	peak date mean		2020/12/27	2020/12/23		2021/04/15	2021/01/11
	N		322	34	4	56	412
Liechtenstein	incidence mean			8654.6			8654.6
	peak date mean			2020/12/18			2020/12/18
	N			1			1
simulation runs	incidence mean	2134.7	3650.4	9159.6		3408.2	3191.5
	peak date mean	2020/03/17	2020/10/22	2020/09/24		2021/06/24	2020/08/17
	N	494	130	70		206	900

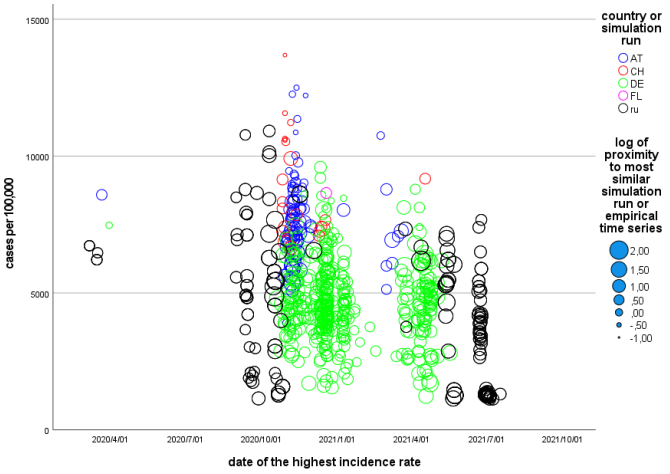


Fig. 4 Distribution of the time of the highest peak of the progression of Covid-19 and the total number of cases in all Austrian, Swiss, Liechtenstein and German NUTS-3 regions and 352 out of 900 simulation runs — circle size is proportional to the logarithm of the proximity between a simulation run and an empirical time series, proximity being measured in the four dimensions of the standardised values of cases and deaths per 100,000, the day of the highest incidence rate and the highest incidence rate that ever occurred; simulation runs are only shown for dissimilarities $< 0.45 \equiv \log \text{proximity} > 0.34$ — all of the unselected runs are simulation runs with only one wave with a peak around 2021-04-18

populated regions is not great) whereas the small red dots represent countries and territories elsewhere in the world with less than 80,000 inhabitants[11] The thick coloured dots represent the simulation runs that fitted the empirical data best (colours show the value of the `BACKGROUND` variable). The diagram shows — not surprisingly — that there is a high correlation between the maximum incidence rate and the final total number[12] of 0.637 over all 900 simulation runs reported, for those used for Figure 4 it is even 0.837 — empirical time series: 0.827). Total death and total infection numbers are also strongly correlated (0.925 / 0.962 /0.860) whereas the correlation between daily new infections and daily death numbers is more modest — mainly because of the delay between infection and death (when the infection leads to death at all), such that it is of minor interest.

This similarity between empirical time series and simulation runs does not only apply to the two output metrics depicted in Figure 3 but also extends to the progress of the epidemic in pairs of NUTS-3 regions and similar simulation runs, as they are shown in Figure 5; for one of these — Figure 5 b comparing

[11] These countries and territories are, with increasing population size, Gibraltar, San Marino, Liechtenstein, Monaco, the Faroe Islands, Bermuda, the Cayman Islands, Andorra and the Isle of Man. Data used stem from the worldometer website, https://www.worldometers.info/coronavirus/, retrieved on September 18, 2020.

[12] As this version of the paper was written long before the end of the epidemic in Germany and elsewhere, at a time when the "fourth wave" had already started, the "final total number" is the number of infected persons on August 26, 2021.

the Haßberge region (in North-East Bavaria) to a simulation run — some input parameters and output metrics are shown in Table 4.

Table 4 Data about the German NUTS-3 region of DE267 Haßberge, Bavaria, output metrics and input parameters of the similar simulation run depicted in Figure 5 b

parameter	Haßberge	Run 958-0425
Population	84,384	38,149
max 7-day incidence per 100,000	32.84	30.33
peak day	472	499
total cases per 100,000	4,939	4,661
total deaths per 100,000	103.1	97.0
case fatality rate	0.0209	0.0210
normative background	—	6
normative drive weight	—	0.6
mutation rate	—	0.05
lockdown start rate	—	100
infections from abroad	—	540.0
men dily moves	—	0.575

4.4 The role of normative behaviour for the progress of a pandemic

From the simulation runs it became clear that the severity of the pandemic does not so much (see Table 2) depend on the normative background but much more so on the weight of the normative drive `ndw`, as the effect of the initialisation of the simulated population with a certain background — determining which risk they prefer to avoid: the risk of loosing their freedom (`BACKGROUND` < 0) or the risk to get infected (`BACKGROUND` > 0) — soon vanishes when the agents

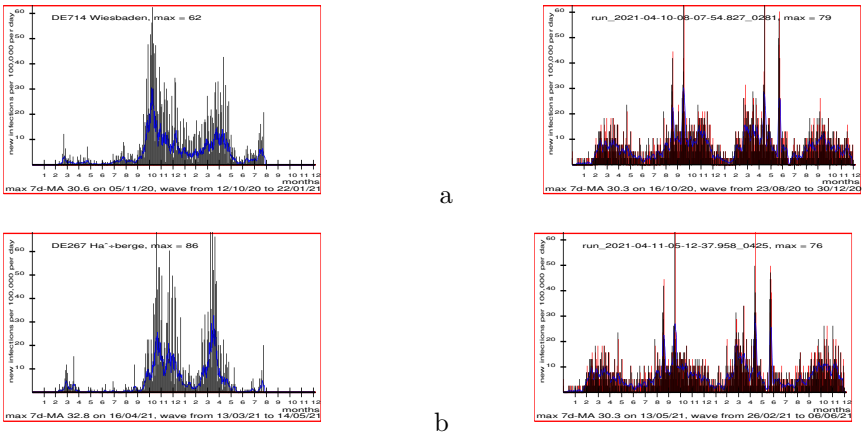


Fig. 5 German NUTS-3 regions compared to simulation runs with a similar pandemic progress. Note: the empirical data are right censored to August 26, 2021, the time of the last download

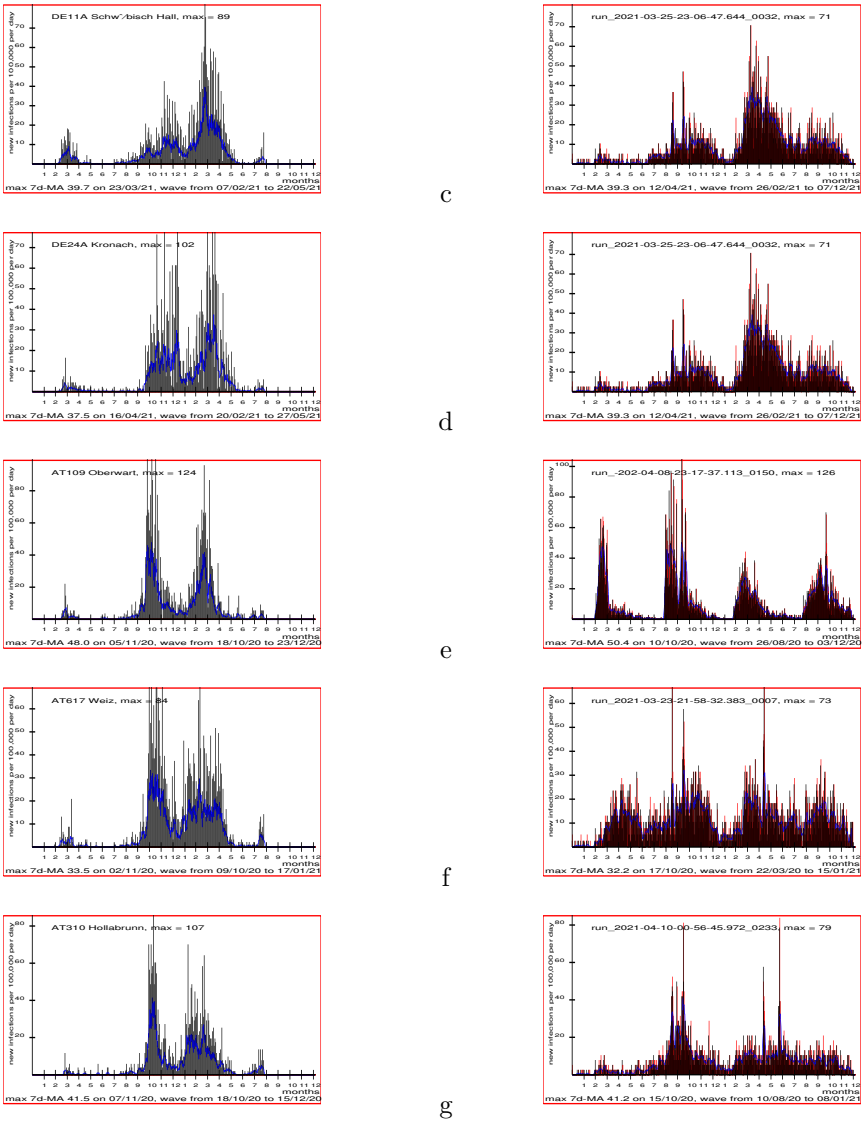


Fig. 6 Continued: German NUTS-3 regions compared to simulation runs with a similar pandemic progress. Note: the empirical data are right censored to August 26, 2021, the time of the last download

exchange information about the behaviour which they apply according to the perceived risk and the normative invocations of others.

The standardised regression coefficients in Table 2 make clear that the progress of the epidemic depends mainly on the mutation rate and on infections from abroad: the higher the mutation rate and the more agents return from abroad infectious the more cases and the more deaths can be expected.

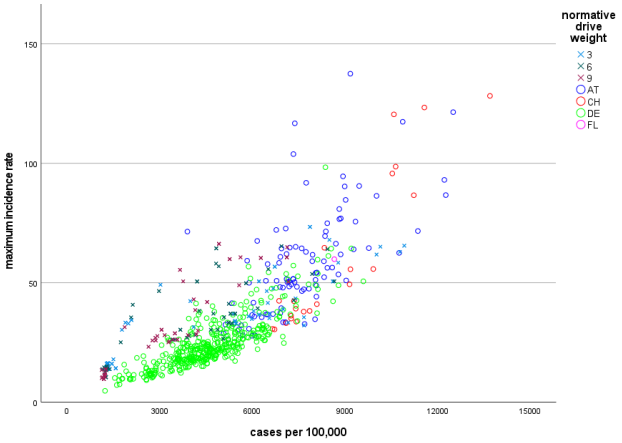


Fig. 7 Comparison between empirical results of German NUTS-3 regions and simulation runs

The population density (as mean distance to nearest neighbour) plays a minor role: if the mean distance is longer slightly fewer cases abd death will occur, but these will occur much earlier (which is only due to the fact that a sparse population distribution leads to an early collapse of the epidemic). With more imported infections the (highest) peak of the epidemic comes later. The normative background with which the agents start into the simulation plays a neglible role, obviously because it is soon superseded by the risk perceived by themselves and the norm invocations received from others — if the agents abide by these invocations (because of a high normative drive weight) the numbers of cases and deaths are smaller than with a low normative drive weight.

The role of lockdown and normative drive weight can be madde more graphic if one calculates the percentage of avoided infections and death as in Table 5.

Table 5 Numbers of cases and deaths averaged over 900 simulation runs for six types of scenarios

	no lockdown		lockdown	
	abs.	%	abs.	%
	cases per 100,000			
ndw = 0.3	4071	100.0	3774	92.7
ndw = 0.6	3518	86.4	3109	76.4
ndw = 0.9	2817	69.2	2299	56.5
	deaths per 100,000			
ndw = 0.3	104.4	100.0	99.0	94.8
ndw = 0.6	92.6	88.7	82.8	79.3
ndw = 0.9	80.2	76.8	65.2	62.5

The estimated positive effect of the highest normative weight as compared to the lowest is about 30 per cent avoided cases and about 23 per cent avoided deaths, with lockdown added another 13 per cent cases and deathss of the

worst case scenario are avoided over the whole range of population density and both mutation rates.

4.5 Other lessons learnt from the model

4.5.1 The role of the settlement structure

Every additional run with identical parameters but with a different `setup-seed` of the random number generator results in a different settlement structure as the centers and boundaries of the `subareas` will be different and as the distribution of the agents over the subeareas will also be different. Although the distribution of the distance between nearest neighbours will be more or less the same for different seeds, the regional distribution of the risk-relevant features of the agents will be different, such that each of these runs produces what can be seen as the history of a pandemic in a different artificial county, canton or departement.

As already mentioned, a different `go-seed` also produces different outcomes, as this seed determines when the agents move and meet an infectious person and when agents are infected as if they had re-entered the simulation region after a stay outside. In the baseline scenario — where there is only one initial infector — only about 10 out of 100 runs with identical parameters and `setup-seed` produced an epidemic as the original infector did not meet any other agent (or sufficiently many agents) which it could infect because all but very few agents it met were too far away, wore masks or just were not hit by the virus the infector was able to spread.

In the more realistic scenarios where infections from outside were possible over the whole run, the current version of the paper concentrates on the combination of one standard `setup-seed` and one `go-seed` such that the settlement structure is always the same and the time when agents virtually returning from abroad are also the same, but whether they are infectious on return can depend on the other input parameters.

4.5.2 Estimating the basic reproduction number R_0

As a side effect of this simulation model it became possible to analyse the validity and reliability of a widely used method to estimate the basic reproduction number R_0 from data about new infections from the past one or two weeks. This method is decribed in detail in an der Heiden and Hamouda (2020) and uses moving averages of recent new infections. The formula used can be written

$$\hat{R_0} = \mathrm{MA}_t / \mathrm{MA}_{t-4} \tag{3}$$

where MA_s is the moving average of the daily number of new infections, either over four or seven days. Originally, R_0 is a theroetical parameter in epidemiological theory. "The reproductive number R_0 [is] defined as the number of secondary infections that arise from a typical primary case in a completely susceptible population. When infection is spreading through a population that may be partially immune, it is often more convenient to work with an effective

reproductive number R, which is defined as the number of secondary infections that arise from a typical primary case." (Wallinga & Lipsitch, 2007, p. 500) Empirically, it is most often impossible to know how many other persons an infected person has infected, and even less possible is to get this number for all persons who were ever infected and infected others within a given period or to find out how many persons, on an average, were infected by all those who are currently infectious. In a simulation model, however, this is possible, as every infectious agent can keep a list of all other agents which it infected during its infectious period. Hence it is possible calculate the mean out of the length of all these lists `whom-I-infected`. This is what the model described here actually does and reports in a plot — see Figure 8. And in parallel to this calculation of the precise reproductive number, it also calculates the two estimates proposed by an der Heiden and Hamouda (2020).

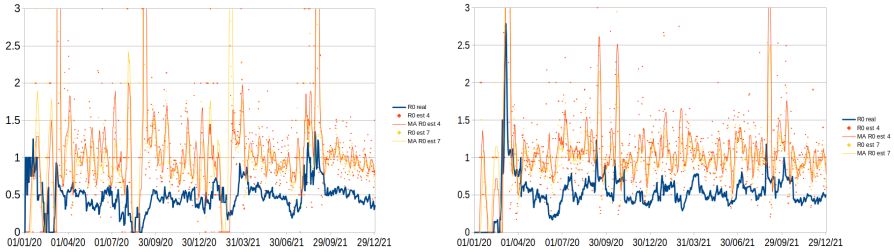


Fig. 8 Comparison between precisely calculated R_0 and its nowcasting estimates in a simulation run resembling an empirical case (left, 38,000 agents, 9.28 per cent ever infected) and a simulation run with a large population and a large number of infected agents (right, 70,000 agents, 6.88 per cent ever infected)

agents	R_0 estimation	mean	std dev
38,000	individual	0.482	0.211
	four days	1.078	1.602
	seven days	1.118	1.607
70,000	individual	0.533	0.274
	four days	1.101	0.912
	seven days	1.048	0.656

Table 6 Comparison of means and variances of the R_0 estimates

Figure 8 shows the difference between the time-dependent mean of the length of all the `whom-I-infected` lists and the estimates on the base of the nowcasting method. Both figures show that the estimated R_0 are much noisier than the R_0 calculated from the number of an infectious agent's individual victims, partly because of the small number of agents contributing to the moving averages (at least in the left diagram, where the exact value, too, shows wider oscillations than the right-hand diagram), but even with a much higher number the R_0 estimated with four-day moving averages is still extremely noisy, and moreover it becomes clear that both estimation methods overestimate the

exactly calculated values: the former are nearly twice as high as the latter — details can be founf in Table 6.

4.5.3 Estimating the dispersion parameter

In recent years it became clear that the reporduction number is not the only parameter determining the progress of an epidemic. Superspreading events showed that the dispersion parameter informing about the overdispersion is at least equally important, as it makes a difference — to give a simple numerical example — whether 10 infectious people infect two others each (resulting in 20 new infections) or whether one out of 10 infectious people infects another 20 while the other nine infect nobody. The distribution of the number of people which infectious people infect is usually modelled as a negative binomial distribution (see for instance Adam et al. (2020); Blumberg, Funk, and Pulliam (2014); Endo, Abbott, Kucharski, and Funk (2020); Lloyd-Smith, Schreiber, Kopp, and Getz (2005)). The literature contains several equivalent definitions of the probability mass function of this distribution. This paper follows a definition of the probability mass function

$$f(k;\ r,p) = \Pr(X = k) = \binom{k+r-1}{r-1}(1-p)^k p^r \tag{4}$$

which contains two parameters p and r which can be estimated from the mean μ and the variance σ^2

$$\mu = \frac{pr}{1-p} \qquad \text{and} \qquad \sigma^2 = \frac{pr}{(1-p)^2} \tag{5}$$

$$p = \frac{\sigma^2 - \mu}{\sigma^2} \qquad \text{and} \qquad r = \frac{\mu^2}{\sigma^2 - \mu} \tag{6}$$

which can be easily calculated from the information the simulation yields about the numbers of agents which were infected by the currently infectious agents. The probability mass function in Lloyd-Smith (2007) — which is in the parameters μ for the mean and k for the dispersion parameter — defines the variance as $\sigma^2 = \mu(1 + \mu/k)$ which leads to $k = \mu^2/(\sigma^2 - \mu)$ which turns out to be the same as r in equation 6 — "decreasing values of k correspond to increasing levels of dispersion" (Lloyd-Smith, 2007). It might perhaps be more desirable to have a parameter $\kappa = \frac{1}{k} + 1, 0 < \kappa \leq N - 2$, whose *increasing* values correspond to *increasing* levels of dispersion[13], and one would avoid $k \to \infty$ for $\sigma^2 = \mu$. In this case $\kappa = 1/k + 1 = 1$ would indicate no overdispersion at all (and the distribution would be the Poisson distribution), and values of $0 < 1/k + 1 < 1$ would indicate the possible case of underdispersion, i.e. the case of $\sigma^2 < \mu$, which, for instance can be observed when the variance is 0 because all infectious persons infected a constant number of susceptible

[13]The maximum possible value of $\kappa = 1/k + 1$ is reached when there is only one infectious person in the population who infected all $N - 1$ others, as in this extreme case the mean is 1 and the population variance is $N - 1$.

persons. This might be an extremely unlikely case, but $0 < \sigma^2 < \mu$ might occur (and it does, in most simulation runs[14]) but cannot be easily detected when only individual infection chains can be observed, as these go back from the person last infected to their infector but do not contain sufficient information about all victims of the same infector. A typical simulation run yields the information that for most of the time about 75 per cent of the infectious agents never infected other agents, which is in line with the observation reported by Endo et al. (2020, p. 3) when they say "High variation in the distribution of secondary cases suggests that most cases do not contribute to the expansion of the epidemic." They estimate k values fromempirical data between 0.04 and 0.6 (which corresponds to $6 < \kappa < 26$ which are typical simulation outcomes — Fig. 9 shows the development of the κ parameter for a simulation run over two simulated years: there are only few short periods with underdispersion, and for most of the time κ is about 3.[15]

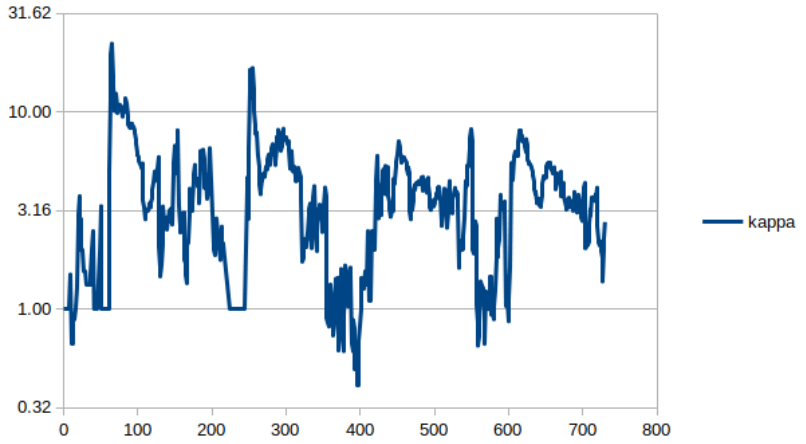


Fig. 9 Development of the *kappa* parameter in the run used in Fig. 8 (left) (logarithmic scale to show the line for $\kappa = 1$)

4.5.4 Estimating the test positivity rate

As another side effect of this simulation model it became also possible to analyse the validity and reliability of the test positivity rate (Wu et al., 2020), as the agents in the model — unlike real persons — know whether they are infected or not. Hence just reading out the values of their instance variable

[14]Consider a small population with currently 77 infectious persons 46 of which never infected anybody else, 23 of which infected one other person, 7 infected two others and one infected three others. This leads to an R_0 of 0.51948 and a variance of 0.51606, yielding $k = -78.961$ and $1/k + 1 = 0.987336$.

[15]A situation with 267 infectious agents of which 181 infected nobody, 64 infected one other agents, 13 infected two, two infected three, one infected four, three imnfected five, two infected six and one infected seven other agents results in $\kappa \approx 3.15$.

`infectious?` yields the test result, and it is possible in the model to differentiate between various testing policies, among them a test of a random sample of the simulated population (which has rarely been done, for instance in Italy (Istituto Nazionale di Statistica, 2020), and Spain, (Pollán et al., 2020) — and in Germany in the near future, (Hoebel et al., 2021) — and often demanded, (Cochran, 2020)) and a test of a non-random sample of agents which are likely to have contact to infectious agents. The latter test strategy is similar to what usually happens when a source of infection was detected (for instance, a wedding ceremony with hundreds of participants some of which showed symptoms in the aftermath of the event, or the well-known case of the bars in a skiing resort in Tyrol, Austria) and most participants and their relatives and neighbours were tested afterwards; the testing of people returning from holidays abroad in risky regions are another example where the probability of finding infected persons is higher than in a random sample.

Wu et al. (2020) describe the problem as follows: "In the absence of systematic random sampling or robust surveillance, true COVID-19 incidence is unknown" and introduce algorithms to provide an "estimate of the true number of infections, which can help not only determine what kind of response is appropriate, but also evaluate the progress or failure of mitigation or containment efforts."

In the model, two ten per cent samples are drawn every day, one entirely random, the other with a selection of all those agents which were near an infectious agent during the day. To be more precise, all agents know how far away they are from an infectious person, and they are selected for the non-random sample in the order of this distance.

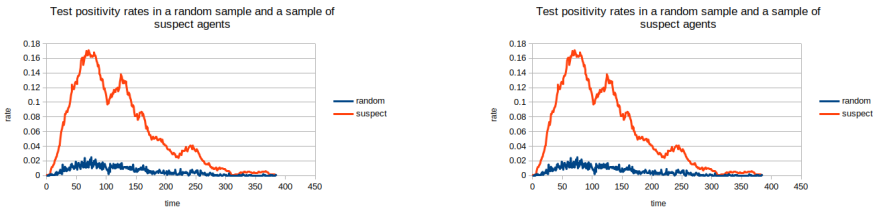


Fig. 10 Comparison between test positivity rate histories in a simulation run resembling an empirical case (top) and a simulation run with a large population and a large number of infected agents (bottom)

It turns out that the test positivity rate for the non-random sample is always higher (which is not a surprise) and even much higher than the test positivity rate for the random sample. The extremely high values which can be observed during part of the simulated epidemic first seemed unrealistically high until Switzerland reported comparable values in its second wave when it published rates of 22.6 and 26.5 per cent for weeks 43 and 44, respectively (Bundesamt für Gesundheit, 2021).

4.6 Vaccination effects

The model can also be used to estimate the effect of a vaccination strategy. Figure 11 shows five simulation outcomes with identical initialisation which differ only in the speed of vaccination doses offered to a population of 30,000 inhabitants.

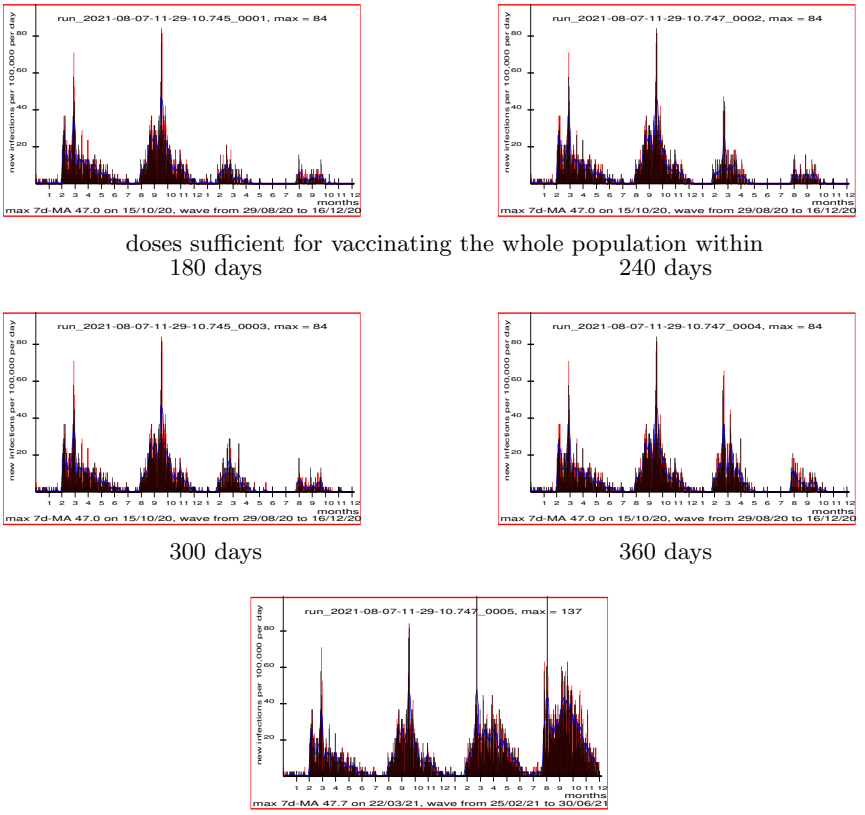


Fig. 11 Comparison between different test strategies starting on December 20 of the first year

As before mid-December of the first year there are no vaccinations at all, the first two waves are the same in all five scenarios. Comparing the first four scenarios with the fifth scenario — the one where vaccination is switched off — shows that vaccination counts, and the more so the earlier a high percentage of the population was vaccinated. As the progress of the vaccination is nearly linear, even a slow vaccination reduces the number of infections considerably as compared to the no-vaccination scenario. The real situation in Germany resembles the fourth scenario where the third wave is nearly as high as in

the no-vaccination scenario and where the fourth wave is much weaker thann waves two and three (as far as this can be judged at the time of writing in September of the second year).

5 Discussion and Conclusion

The model and its various scenarios as well as the empirical data about the progress of the Covid-19 pandemic in all NUTS-3 regions of Germany, Austria and Switzerland showed that the progress of a pandemic can depend on the settlement structure and on the density of contacts with adjacent regions as well on political measures such as lockdowns and vaccination and on the salience of social norms in a (local) society. The research questions listed at the beginning of the paper can now be answered as follows:

norm orientation: Does it make a difference when people in a community are more or less aware of others' behaviour with respect to risk avoidance and shape their own behaviour according to what they observe in their neighbourhood? And does this norm orientation change over time, and if so, how will it change under the risk of an epidemic?

The effect of the background norm orientation at the outset of the pandemic is more or less negligible, but the normative drive weight plays a role: if the agents plan the measures they take according to the norm invocations they received and not only according their private view of the risks, these private measures reduce the overall risk of high incidence rates and high numbers of deaths. The analysis of runs with different values on the normative drive weight showed (see Table 5) that abiding by exchanged norm invocations reduces infection cases and deaths by up to 40 per cent.

settlement structure: Does settlement structure make a difference, i.e. do different allocations of agents to subregions of the simulated world lead to different pandemic histories?

Different settlement structures from different initialisations of the location of villages, village centers and agents lead to different paths through the pandemic, sometimes with considerable effects — some settlement structures result in an early end of the pandemic, but when the pandemic continues the results are similar.

heterogeneity of risk aversion: Does it make a difference when people have different degrees of conviviality and carelessness which make the avoid or seek risky situations?

Results show that different degrees of conviviality and carelessness result in different probabilities of being infected. The model shows this with the difference of the distribution of the product of these two individual constants for agents that were never infected and those that were infected and, perhaps, died. As no empirical data are available to compare this result to, this effect cannot be validated. But the simulation

results are in line with the maximum incidence rates in different age ans sex groups in Germany (see Table B1).

person-zero assumption: Can current pandemic histories be explained with the assumption that one infected person intrudes a community?

In most runs with only one agent initlally infectious the pandemic ends after a first wave and remains local — an impression one could have had in early 2020 when outbreaks of Covid-19 were due to local superspreading events, but soon it became clear that the pandemic was quickly exported from these small regions and intruded into adjacent regions and regions far away. Hence it was necessary to endow the model with the fiction of agents returning from abroad, bringing new infectiosity in the simulated region which was otherwise closed.

mutation: Can current pandemic histories be explained with the assumptions that mutations do not affect infectiosity?

Runs with mutation and without mutations did not differ much, qualitatively, but the final numbers of cases and death were considerably higher in runs with mutation.

reliability of estimates of R_0: Is the estimate of the basic reproduction number R_0 with the nowcasting method an der Heiden and Hamouda (2020) reliable, and if so, under which conditions?

The simulation showed that R_0 was always overestimated by the nowcasting method — and R_0 values were rarely published after the first few months of the pandemic in Germany after the first few months of the pandemic. This seems to be due to the fact that overdispersion was rather the rule than the exception: Few infectious agents were responsible for many infections, and many agents never infected others. This seems also to have been the case in many countries.

reliability of estimates of the percentage of positive tests: Is the percentage of positive tests on all tests (the test positivity rate, see Wu et al. (2020)) a reliable measure of the severity of a pandemic?

Without a random sample of persons to be tested test positivity rates do not seem to be reliable. Hence incidence rates seem to have been underestimated in most countries.

vaccination: What is the effect of a late start of vaccinations and of the availability of doses, i.e. how long it takes until the total population can have had an opportunity to be vaccinated?

A vaccination effect could be found in the respective simulation — whether this effect could be reliably explained and replicated in real countries and regions remains an open question: either states succeeded in vaccinating growing rates of all their citizens, or they failed, but in the latter case all other parameters of the real epidemic in these countries are not very likely to be reliable.

Some of these research questions could not be answered completely, but with a pandemic still continuing this coild be expected. Further extensions of the model — for example the inclusion of variables such as age and sex (with

different distributions of conviviality and risk aversion and different activity ranges — might be desirable.

6 Appendices

Supplementary material. Additional tables are available on request as supplementary material.

Acknowledgments. The algorithm for the calculation of the salience of norms was developed in the EMIL ("Emergence in the Loop") project supported by the European Commission in the Sixth Framework, contract number 033841. Thanks go to Marc Hannappel and Ulf Lotzmann who ran part of my simulations on their (more powerful) machines.

Data availability statement. The empirical date used in this paper were downloaded from Bundesamt für Gesundheit (2021); Bundesministerium für Soziales (2021); Robert-Koch-Institut (2021). The aggregated data for NUTS-3 regions and days (D-A-CH.xlsx) are available from https://doi.org/10.25937/7vkh-tt08.

Conflict of interests. There are no competing interests. This paper did not receive any support from anybody except my colleagues who ran part of my simulations on their machines.

Code availability statement. The NetLogo model is available at ComSES at https://doi.org/10.25937/7vkh-tt08

Compliance with ethical standards. No ethical standards whatsoever can be violated by the research documented in this paper.

Appendix A Minding Norms in a Nutshell

The model presented in this paper uses the theory developed in the EU projects EMIL and GLODERS and explicated in in Conte et al. (2013) and Elsenbroich et al. (2016) which partly go back to Cialdini, Reno, and Kallgren (1990). The main idea behind this theory is that humans can be modelled as using an architecture (called the "EMIL-I-A architecture" (Conte et al., 2013, p. 162–166)) in which norm invocations, compliances, violations, sanctions and punishments are stored. To this end, the agents use a memory in which for each norm the number of

- their own compliances and violations (C and V, respectively),
- compliances and violations observed with neighbours (O_C and O_V, respectively),
- sanctions and punishments received (S and P, respectively) and
- explicit norm invocations (when agent A observes a compliance of violation of a norm by agent B and communicates this to the latter for instance in the form of praise or blame, E_C and E_V, respectively)

is stored. These numbers are decremented by a certain percentage now and then and computed to generate the *salience* of the respective norm according to the following formula:

$$\sigma = \alpha \left(\beta + \frac{C - V}{C + V} w_c + \frac{O_c - O_v}{O_c + O_v} w_o + \frac{\max(0, (O_v + V) - P - S)}{O_v + V} w_{npv} + \frac{P w_p + S w_s}{\max(P + S, O_v + V)} + \frac{E_c - E_v}{E_c + E_v} w_e \right) \tag{A1}$$

where the capital letters have the meaning explained above. The coefficients w_c, w_o, w_{npv}, w_p, w_s and w_e are the weights for the six factors ("norm cues") derived from Cialdini et al. (1990) and defined in Andrighetto et al. (2013) (see also Andrighetto and Castelfranchi (2013) and Troitzsch (2018)). α and β have to be chosen dependent on the weights w_c, w_o, w_{npv}, w_p, w_s and w_e in a way that $0 \leq \sigma \leq 1$. The weights used in the literature cited above are $w_c = w_s = w_e = 0.99, w_{npv} = 0.66, w_o = w_p = 0.33$.

The model described in this paper uses only C, V, O_C and O_V. The other four kinds of "norm cues" could be added in future extensions of the model.

Appendix B Additionall tables

Table B1 Maximum incidence rates for age and sex groups in Germany from January 2020 through August 2021

age group	male	female
0 .. 4	26.8	26.6
5 .. 14	46.4	45.9
15 .. 34	49.1	49.6
35 .. 59	41.1	51.9
60 .. 79	33.5	32.1
80+[a]	59.3	82.6

[a]That the oldest group shows the highest incidence rate is obviously due to the fact that these persons have a weaker immune system and that the proportion of those living in precarious environment such as nursing homes for the elderly is high.

Table B2 Input parameters kept constant over all experiments documented in this paper

input parameter	value
asymptomatic-per-infected	0.65
case-fatality-rate	0.04
far-commuter	0.65
go-seed	-1637705880
heterogeneity	wrt risk
illness-duration-max	30
illness-duration-min	12
immunity-duration-max	30
immunity-duration-min	24
infection-distribution	beta 5 2
incubation-time-max	20
incubation-time-min	2
infectiosity-duration	24
initial-subareas	13
interruptible?	FALSE
long-distance	80
reporting-delay	0
reporting-rate	1.0
superspread?	FALSE
tested-per-infected	1.0
Time-till-immunity-max	2
Time-till-immunity-min	0
Time-to-infectiosity-min	0
travellers	0.015

Table B3 Input parameters and output metrics of the runs documented in the Figures

	Fig. 1 024_0089	Fig. 2 214_0027	Fig. 5 a 827_0281	Fig. 5 b 958_0425	Fig. 6 cd 644_0032	Fig. 6 e 133_0150	Fig. 6 f 383_0007	Fig. 6 g 972_0233	Fig. 11 180 745_0001	Fig. 11 240 747_0002	Fig. 11 300 745_0003	Fig. 11 360 747_0004	Fig. 11 no 747_0005
Input parameters													
auto-lockdown	off		internal	external	external	internal	external	internal			off		
BACKGROUND	2	0	6	6	3	-6	-3	0			6		
constant-seed	setup-constant				all-constant						all-constant		
distance-for-infection	1.33	0.5			0.5						0.5		
infectious-from-abroad	none	seasonal			seasonal						holidays		
lockdown-at	100		100	100	100	50	100	100			100		
lockdown-end-at	55	55	35	35	35	17.5	35	35			70		
max-months	9	24			24						24		
move-speed	0.27	2.4			2.4						2.4		
mutation-rate	0	0.05			0.05						0.05		
NDW	0.75	0.6	0.6	0.6	0.6	0.9	0.3	0.6			0.3		
setup-seed	-2078345452	-1941496835			-1941496835						-1941496835		
short-distance	3.3	8			8						8		
tested-per-infected	0.45	0.05			0.05						0.05		
time-to-infectiosity-max	3	4			4						4		
total-population	38149	38149	38149	38149	38149	70000	38149	38149			38000		
zero-person	densest	sparsest			sparsest						sparsest		
vacc-lasts-days	N/A				N/A				180	240	300	360	N/A
vaccination?	FALSE				FALSE				TRUE	TRUE	TRUE	TRUE	FALSE
%continuous	35	20			20						20		
Output metrics													
still susceptible	37968	36286	36248	36334	35743	64858	35157	36105	7810	7766	7808	7745	34472
infected	0	79	32	43	12	56	34	16	1	1	1	2	40
ill	0	31	11	21	7	25	15	11	0	0	1	0	31
recovered	2	24	7	14	11	19	13	3	0	0	0	0	17
immune	178	1690	1811	1700	2326	4917	2866	1969	30148	30190	30148	30204	3352
dead	1	39	40	37	50	125	64	45	41	43	42	49	88
infectious	0	108	44	62	23	84	47	27	1	1	2	2	72
never met infectious	37814	35285	35169	35279	34336	62404	34232	34972	7714	7648	7683	7603	33540
escaped infection	154	1001	1079	1055	1407	2454	925	1133	96	118	125	142	932
internally infected	180	1619	1686	1609	2206	4128	2803	1830	29802	29838	29792	29836	2976
returned infected	1	244	215	206	200	1014	189	214	388	396	400	419	552
no more infectious at	4272	5040	-1	-1	-1	576	-1	-1	-1	-1	-1	-1	-1
median co ca su	0.2162	0.2156	0.2157	0.2156	0.2138	0.2166	0.2129	0.2154	0.3962	0.3957	0.3946	0.3939	0.2116
median co ca in	0.3355	0.3004	0.3023	0.3113	0.3148	0.2583	0.3079	0.3048	0.1887	0.1890	0.1891	0.1897	0.3037
wear a mask	16256	12236	11692	11452	11619	31244	5656	11439	5643	5730	5747	5840	5827
stay at home	16461	12302	11680	11497	11425	31220	5577	12265	5797	5807	5865	5844	5878
mean distance	0.2327	0.2325	0.2325	0.2325	0.2325	0.1682	0.2325	0.2325	0.2318	0.2318	0.2318	0.2318	0.2318
sdev distance	0.1672	0.1723	0.1723	0.1723	0.1723	0.1294	0.1723	0.1723	0.1728	0.1728	0.1728	0.1728	0.1728
when-max-infected	88	370	293	502	471	287	294	292	292	292	292	292	450
n-of-max-infected	6.1429	12.2857	11.5714	11.5714	15.0000	35.2857	12.2857	15.7143	17.8571	17.8571	17.8571	17.8571	18.1429
when-max-deceased	115	58	280	515	551	679	334	279	307.0000	307.0000	307.0000	307.0000	307.0000
n-of-max-deceased	0.1429	0.5714	0.5714	0.4286	0.7143	1.5714	0.5714	0.8571	1.0000	1.0000	1.0000	1.0000	1.0000
mean-daily-moves	0.0610	0.5737	0.5719	0.5749	0.5805	0.5723	0.5846	0.5750	0.5772	0.5886	0.5779	0.5979	0.5781
vaccinated	0	0	0	0	0	0	0	0	28728	28650	28646	28515	0
vaccination-refused	0	0	0	0	0	0	0	0	7937	7932	7924	7886	0

Appendix C ODD+D Description of the Model

This section describes the model along the lines of ODD+D (Grimm et al., 2006, 2010; Müller et al., 2013).

A. **Overview**

I. ***Purpose***

I.i. *What is the purpose of the study?* The purpose of this study is to find out how the course of an epidemic is determined by the settlement structure of a population and of certain behavioural features, among them the propensity to abide by norms invocated by others or to violate them. Additionally, the model allows the judgement of the quality of the results of empirically estimating parameters such as the basic reproduction number R_0 and the dispersion parameter κ (see C.I.ii.8.) as well as of the test positivity rate (see C.I.ii.9.). Finally, a simulation is possible of the effect of vaccinations depending on when it starts and how fast sufficient numbers of doses are available (see C.IV.ii.34.).

I.ii. *For whom is the model designed?* For researchers and students of epidemiology as well as for social scientists at large.

II. ***Entities, state variables and scales***

II.i. *What kinds of entities are in the model?* There is only one class of active entities in the model, namely the `people`. For technical reasons, there is also the entity type `subarea` which is used to allow for different population densities all over the simulated world; `subareas` also collect information on all `people` in the `ill` state.

II.ii. *By what attributes (i.e. state variables and parameters are these entities characterised?*

- `people` keep a number of initialised constants:
  - `conviviality` and
  - `carelessness`

  which determine their persional risk aversion, and of instance variables
  - `state`: suspensive, infected, ill, recovered, immune, dead,
  - `infectious?`: for some of the infected, all ill and some of the recovered person agents (infectiosity starts only some time after being infected and ends some time after recovering),
  - their place in the topography of the simulated world,
  - `wears-a-mask?`: and
  - `stays-at-home?`: modifying the probability of being infected by another agent in its neighbourhood,
  - `perceived-infection-risk`: the current risk of being infected as the proportion of infectious neighbours among all neighbours (in times of increasing daily new infcetions) or as the proportion

of all infectious agents among all agents (in times of decreasing daily new infections),
  - — in the normative mode — the salience of four norms and the propensity to abide by them or to violate them (see the discussion in C.IV.i.4.).
- `subareas` keep a number of initialised constants (area, population, colour) andstate variables for counting their people with respect to their state. These are output at the end of each run for further analysis.

II.iii. *What are the exogenous drivers of the model?* There are no exogenous drivers of the model except the possibility that a user can interrupt a run and set or unset the global variable `lockdown?` which reduces the mobility of all agents (see below C.IV.ii.13.–C.IV.ii.15.). The chooser `infectious-from-abroad` is not an exogenous driver proper, as it is set before the start of the model, but has an effect similar to the effect of an exogenous driver (see below C.IV.ii.6.). The switch `superspread?` is not an exogenous driver either, as it is also set before the start of the model, but has also an effect similar to the effect of an exogenous driver (see below C.IV.ii.7.).

II.iv. *If applicable, how is space included in the model?* Space plays a prominent role in the model as the distribution of agents over the simulated world differs between different random initialisations.

III. ***Process overview and scheduling***

III.i. *What entity does what, and in what order?* The only active entities are the `people` agents which are scheduled at the start of the simulation to move to another patch. The time when they will move is random. When the move is executed they will not only move but schedule their next move at a random time in the future, and if they meet an infectious agent at or near the new patch they get infected, get the time scheduled when they become infectious and ill, respectively, and when they fall ill, they are moved to a central patch in their `subarea` (the "hospital") where they stay until the scheduled time for recovery or death. In both cases they return to their start patch ("`home`"), and the loss of infectiosity as well as the start and end of their immunity period are scheduled for those which recovered. The time between movements is hard coded as a uniformly distributed random number of 0 to 4 simulated days, whereas the uniform distributions of the other period lengths can be set via the GUI (see below C.IV.ii.22.–C.IV.ii.26.).

B. **Design Concepts**

I. ***Theoretical and Empirical Background***

I.i. *Which general concepts, theories or hypotheses are underlying the model's design at the system level or at the level(s) of the submodel(s) . . . ?* The model draws on standard epidemiological models

and extends them in two directions as it wants to give answers to the questions:

I.i.1 What is the consequence of different settlement structures and of the assumption that every person is able to infect any other person regardless of the distance between them (see **??**)?

I.i.2 What is the consequence of the violation of the assumption that a population is homogeneous with respect to the personal risk aversion of its members (see C.IV.ii.33.)?

I.i.3 Does communication among people about infection risks and measures taken against infection make a difference (see C.IV.i.4. and C.IV.ii.33.)?

I.ii. *On what assumptions are the agents' decision models based?* In the non-normative version of the model, agents do not make decisions proper but just move around and are exposed to the risk of infection and of falling ill. Only in the normative version, they learn from others how they should behave and make a normative decision of taking one of the actions of putting a mask on or off or of staying at home whenever possible or moving around without any necessity. Their decisions are controlled by the personal salience of the four norms to take one of the four actions. For more details see Andrighetto et al. (2013); Andrighetto and Castelfranchi (2013); Troitzsch (2018) and Section A where the formula for the calculation of the salience of a norm together with the values of the weights of the memory contents are given. The term β in this formula is represented by the variable `beta` in the NetLogo code, whereas the variable `one-divided-by-alpha` in the code correspond to the inverse of the term α of the formula. The calculation of `one-divided-by-alpha` and `beta` in the code of the procedure `calculate-a-salience` makes sure that, with any values of the weights w_c, w_o, w_{npv}, w_p, w_s and w_e of the formula, the resulting value of a salience is always $0 \leq \sigma \leq 1$.

I.iii. *Why are certain decision models chosen?* See again Andrighetto et al. (2013); Andrighetto and Castelfranchi (2013); Troitzsch (2018).

I.iv. *If the model ... is based on empirical data, where does the data come from?* The parameters of the model are taken from the current discussion about the parameters of the Covid-19 pandemic. As such they are currently not reliable and save only as a starting point for sensitivity analysis and calibration of the non-epidemiological input parameters.

I.v. *At which level of aggregation were the data available?* Only at the aggregate level, as information about the length of periods between infection, infectiosity, illness, recovery and death are still only roughly estimated, and the estimates differ from study to study.

II. ***Individual Decision Making***

II.i. *What are the subjects and objects of decision making? On which level of aggregation is decision making modelled?* Objects of decision making in the normative mode are the actions of putting a mask on or off and of staying at home or of leaving one's home even if unnecessary.

II.ii. *What is the basic rationality behind agents' decision making in the model? Do agents pursue an explicit objective or have other success criteria?* There is a certain rationality behind agents' decision making in the normative mode, namely to behave like the neighbours expect them to do.

II.iii. *How do agents make their decisions?* Both by estimating an objective risk of being infected and by deriving a propensity of action from the current saliences of norms which in turn depend on the contents of their memories of earlier norm compliances, violations, invocations, sanctions or punishments. For details see Andrighetto et al. (2013); Andrighetto and Castelfranchi (2013); Troitzsch (2018) and Section A.

II.iv. *Do the agents adapt their behaviour to changing endogenous and exogenous state variables?* Yes, but only in the normative mode. In the non-normative mode it is the probability of being infected that determines their future fate.

II.v. *Do social norms or cultural values play a role in the decision making process?* Yes, in the normative mode, explained in the previous items.

II.vi. *Do spatial aspects play a role in the decision process?* Yes, norm saliences depend only on invocations from near neighbours.

II.vii. *Do temporal aspects play a role in the decision process?* No.

II.viii. *To which extent and how is uncertainty included in the agents' decision rules?* The decision making process ends in a propensity to act this way or that way, and this propensity is taken as the probability with which an action is taken.

III. ***Learning***

III.i. *Is individual learning included in the decision process?* Only in terms of norm learning, see above.

III.ii. *Is collective learning implemented in the model?* No.

IV. ***Individual Sensing***

IV.i. *What endogenous and exogenous state variables are individuals assumed to sense and consider in their decisions? Is the sensing process erroneous?* Agents sense which neighbour is infectious (from which they calculate their personal risk which is only used as the probability that they are infected). Risk perception does not affect the decision where to move next.

IV.ii. *What state variables of which other individuals can an individual perceive? Is the sensing process erroneous?* The only state variable of this kind is `infectious?`. Norm invocations are transmitted as simple messages written directly into the memory of the recipients.

IV.iii. *What is the spatial scale of sensing?* Sensing is only possible in the neighbourhood which is defined by `distance-for-infection`; other information can come from within a radius of five patches.

IV.iv. *Are the mechanisms by which agents obtain information modelled explicitly, or are individuals simply assumed to know these variables?* As for infectiosity, they just know, norm invocations are transmitted in simple messages.

IV.v. *Are costs for cognition and costs for gathering information included in the model?* No.

V. ***Individual Prediction*** There are no predictions.

VI. ***Interaction***

VI.i. *Are interactions among agents and entities assumed as direct or indirect?* Direct.

VI.ii. *On what do the interactions depend?* Interactions depend on vicinity.

VI.iii. *If the interactions involve communication, how are such communications represented?* The sender of a norm invocation writes directly into the memory of the recipient agent.

VI.iv. *If a coordination network exists, how does it affect the agent behaviour? Is the structure of the network imposed or emergent?* There is no network.

VII. ***Collectives*** There are no collectives or other aggregations.

VIII. ***Heterogeneity*** Both `people` and `subarea` agents are homogeneous as they have the same structure, processes are equal among them, but some state variables are randomly assigned during the initialisation and change over time. Heterogeneity with respect of the initial values of these state variables can be switched on and off (see C.IV.ii.33.).

IX. ***Stochasticity*** Periods between events are stochastic, actions are taken according to a propensity which is turned into a probability to take this action.

X. ***Observation***

X.i. *What data are collected from the ABM for testing, understanding and analysing it, and how and when are they collected?* Two agents (`witnesses`) write whatever they do into a log file, at the end all current information about the state of the model and all its agents is stored for inspection. The logfile contains some additional information about distribution parameters at the end of the run.

X.ii. *What key results, outputs or characteristics of the model are emerging from the individuals? (Emergence)* As there is no emerging structure beside the frequency distributions of the state of the people, nothing important can be observed in this respect. The form of the time series of daily infections and death can, however, be seen as an emergent phenomenon, as this form deviates considerably from the results of the classical SIR compartment models.

C. **Details**

I. ***Implementation Details***

I.i. *How has the model been implemented?* The model is implemented with NetLogo 6.2.0. The buttons, sliders, switches etc. with which global parameters (see below C.IV.ii.4.) have the following meaning (from top left to bottom right, left of the view):

I.i.1 `setup` and `go` have the usual meaning, but in an event-oriented model `go` is a *go forever* and can either be stopped with `Tools->Halt` (but then at the risk of loss of all final output except the logging file and the PNG files of the view output every simulated day) or via an interruption mechanism explained below (see C.IV.ii.32.).

I.i.2 `Date and time` gives the simulated time which starts at the beginning of the year 2020.

I.i.3 `stop date` shows how long the simulation can run (namely for `max-months` months, if it is not interrupted before or ends when no infectious agents are left.

I.ii. The monitors and plots are more or less self-explanatory:

I.ii.1 (left of the view, these and the following plots have little monitors attached that inform about the median and maximum of the respective distributions) the two plots `conviv*careless...` show the distribution and the median of the product of the two individual constant attributes `conviviality` and `carelessness` separately for the susceptible and for all other agents and the individual variable attribute `perceived-risk`.

I.ii.2 the plot `conviv*carel*perc...` shows the distribution and the median of the product of the two individual constant attributes `conviviality` and `carelessness` and the individual variable attribute `perceived-risk` which is at the same time the probability of being infected once an infectious contact happened.

I.ii.3 `perc inf risk` shows the distribution and the mean of the individual variable attribute `perceived-risk` which while the moving-average of the daily number of newly infected agents rises (`pandemic-grows?` is true) is the percentage of infectious agents within a radius of 5, otherwise it is the overall percentage of infectious agents.

I.ii.4 `med perc risk` represents the maximum and the median of the distribution of the `perceived-risk` attribute over time.

I.ii.5 (right of the view) `current state of the population`: these two plots represent several global percentages,

- the upper plot containing the number of susceptible agents and of all who were ever infected (per 1,000),
- the second plot containing the numbers of currently infected, infectious, ill, recovered, immune and deceased persons (per 1,000),

separated for better legibility.

I.ii.6 `infectious in last 24 hours` represents the number of newly infected, those whose test was reported and a 7-day moving average of the former (per 1,000).

I.ii.7 `deaths in past 24 hours` represents the number of newly deceased, those which were reported and a moving average of the former (per 1,000).

I.ii.8 The `R0 real / est` plot keeps track of three variants of the *basic reproduction number* R_0; the "real" reproduction number is calculated as the mean of the length of the list `whom-I-infected` of all currently infectious agents, the two others are calculated according to the rules published by the German Robert Koch Institute, in two variants, one using 4-day moving averages, the other uses 7-day moving averages of daily new infections, R_0 being calculated as the quotient of the current moving average and the moving average four days earlier, hence $\mathrm{MA}_t/\mathrm{MA}_{t-4}$. See an der Heiden and Hamouda (2020). Moreover, it keeps track of a dispersion parameter which for this model is defined as $0 \leq (\sigma^2-\mu)/\mu^2+1 \leq N-2$ and called $\kappa = 1/k + 1$ with k as in the formula for a negative binomial distribution of Lloyd-Smith (2007). The dispersion parameter defined here is 0 when there is no variance at all (all infectors infected the same number of agents), it is 1 when there is neither underdispersion nor overdispersion ($\sigma^2 = \mu$ as in the Poisson distribution), and it is $N-2$ when only one agent infected all others in the population.

I.ii.9 The `test positivity rate` plot keeps track of two variants of the *test positivity ratio*, one calculated from a five per cent random sample of the overall population and one calculated from a sample of equal size containing all infectious agents and their nearest neighbours.

I.ii.10 The `whom-I-infected_t` plot tries to visualise the time dependent distribution of the number of the infections directly caused by the currently infectious agents.

I.ii.11 (below the view) `max daily infected` and `max daily dead` contain two numbers: the number of the day when the maximum of daily new infections and deaths, respectively, was reached, and the number of these events.

I.ii.12 `incidence rate` the number of new infections per 100,000 during the past seven days[16].

I.iii. In each run, the model creates a directory into which four files are written (the final state of the world and a profile[17] — except in behaviour space —, a logging file with details about interesting

[16] Germany, Switzerland and Austria use this definition for the incidence rate, others use the number of infections per 100,000 or per million and/or during the past fourteen days and call it *14-day notification rate*, e.g. in https://www.ecdc.europa.eu/en/publications-data/weekly-subnational-14-day-notification-rate-covid-19 with fourteen days per 100,000.

[17] This is a means of finding out how time-consuming the procedures of the model are.

events, a result file with details about the epidemic histories in the individual subareas, and the data generated for all plots). At the end of every simulated day — except in behaviour space —, a PNG file is written, and these PNG files can be converted to an AVI file using (on Unix systems) `mencoder` (NetLogo's movie feature is not used as not all video programmes can work with MOV files).

Which events are logged can be determined in the code by changing the first parameter of the util-flog procedure and/or by setting the respective switch.

II. ***Initialisation***

II.i. *What is the initial state of the model world, i.e. at time $t = 0$ of a simulation run?* During model setup global variables (mainly counters and file names for output) are initialised and the entities are generated according to the input parameters. `people` agents are generated by the `subareas` and placed on patches belonging to each `subarea` according to the settings of the global parameter `total-population` (see C.IV.ii.5.); `people` agents get their normative memories initialised according to C.IV.i.4., but to which extent they use the information in their memories depends on the parameter `NDW`, the weight of the normative drive which can be reduced to 0.

II.ii. *Is the initialisation always the same, or is it allowed to vary among simulations?* The initialisation is equal for equal random number generator seeds. This seed is different for each run of the model provided that the global parameter `constant-seed?` is switched off.

III. ***Input Data***

III.i. *Does the model use input from external sources such as data files or other models to represent processes that change over time?* No.

IV. ***Submodels***

IV.i. *What, in detail, are the submodels that represent processes listed in 'Process overview and scheduling'?* Four submodels can be identified (although they are not explicitly separated in the code as this is not possible in NetLogo):

IV.i.1 A calendar makes sure that at the end of a simulated day all changes in the epidemiological state of the population are collected and sent to the respective plots and monitors of the user interface. This also includes an update done by all `people` with respect to their actual neighbourhood.

IV.i.2 `people` schedule their own movements around their `home` which was assigned to them and to which they return every simulated Saturday. These movements happen at discrete events, they are scheduled at the time of current movement for some random time in the future. The following events can be scheduled by the model (for details see the flowchart in Figure C1):

- `move-around`: the agent moves to another place,
- `set-infectious`: the agent is infected and becomes infectious,

- `report-a-positive-test`: the agents is reported to the observer as infected,
- `fall-ill`: the agent shows symptoms and is brought to hospital,
- `recover-or-die`: the agent either recovers or dies and is returned to its original patch,
- `report-a-death`: the death of the agent is reported to the observer,
- `become-immune`: the agent is no longer infectious nor susceptible,
- `lose-immunity`: the agent loses its immunity and is susceptible again.

IV.i.3 When as a consequence of their movements they approach other `people` agents which are infectious they take the risk of being infected. This risk depends on the values of their `conviviality` and `carelessness` instance variables (which are initialised to constant values at the simulation start and remain constant) and on their currently `perceived-risk` as well as — in the normative mode — on their earlier decision to wear a mask and to stay at home (stored in the instance variables `wears-a-mask?` and `stays-at-home?`). The event of being infected happens immediately, the events of becoming infectious, of having first symptoms, of having to be hospitalised, of recovering, becoming immune or losing immunity as well as death are scheduled for random times in the future whose distribution is a uniform distribution with the parameters noted below in C.IV.ii.22.–C.IV.ii.26..

IV.i.4 In the normative mode, `people` process the norm invocations they have stored in their memories, calculate their actual saliences (see Section A) and decide whether they change their behaviour with respect to wearing a mask and staying at home[18] and whether they let their neighbourhood know about their decision, encouraging them to behave the same.

IV.ii. *What are the model parameters, their dimension and reference values?* The following list informs about all the global parameters of the model (standard values in paranthese, multiple values used for the sensitivity analysis and calibration, those in boldface hold for the comparison of vaccination strategies):

IV.ii.1 `total-population` determines the number of `people` in the simulated world. Usually there will be several agents per patch in the initialisation; the algorithm in the setup procedure makes sure that each subarea has the same number of inhabitants, hence its density is indirectly proportional to its area, and that the total

[18] In a real world scenario this would mean: wearing a mask whenever near other people and staying at home except for going to work or meeting a doctor or to do the most necessary shopping as in mild phases of lockdown.

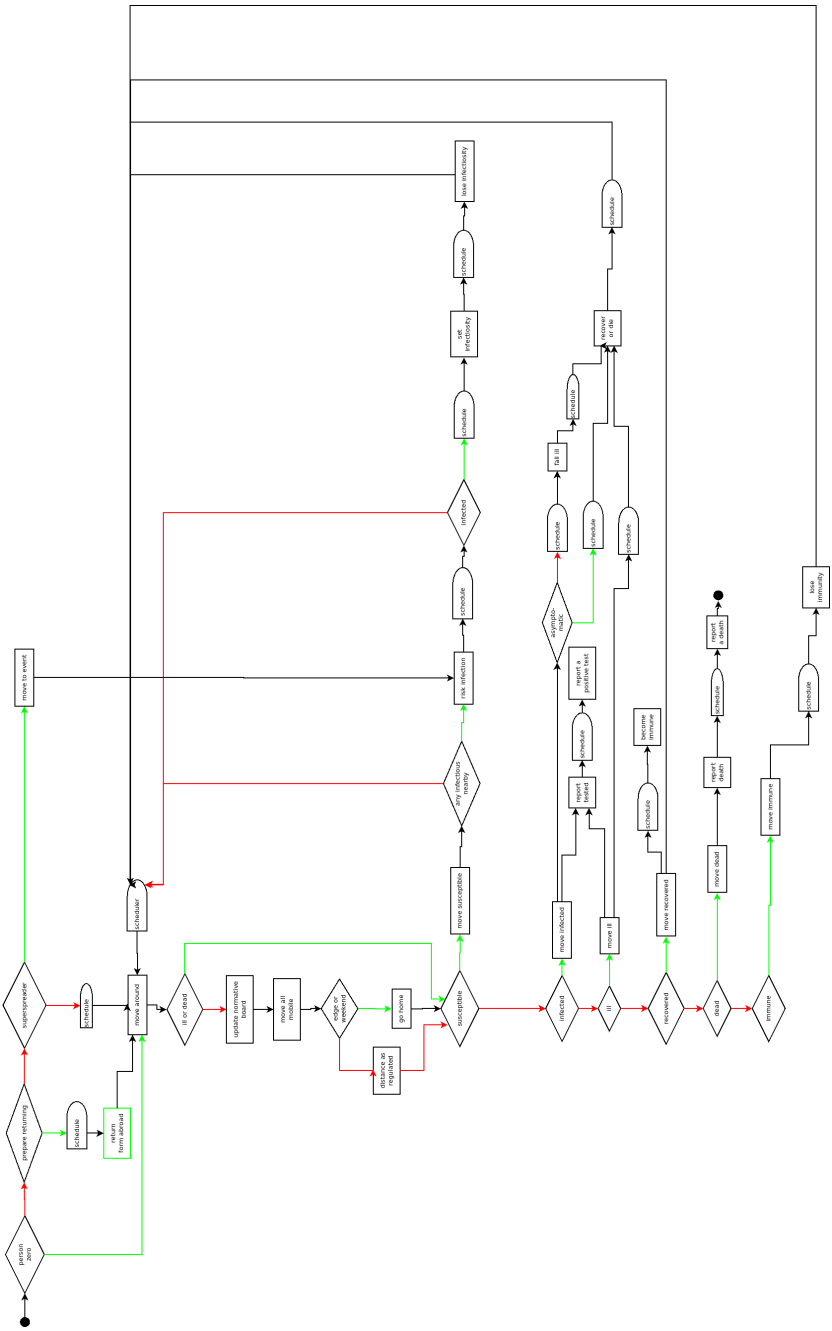


Fig. C1 Flowchart of the events in the model. Legend for the edges: green if the condition is true, red if the condition is false

population is as determined with this input parameter (**38,000**, 70,000).

IV.ii.2 `BACKGROUND` sets the normative predisposition, i.e. the number of norm invocation which agents have in their memories when the simulation starts (-6, 0, **6**). With `heterogeneity` set to `none` or `wrt risk`, all agents have the same number of norm invocation in their memories. `BACKGROUND > 0` sets the initial number of invocations of risk-aversive norms to `BACKROUND`, `BACKGROUND < 0` sets it to `-BACKGROUND`. With `heterogeneity` set to `wrt norm` or `both`, the initial number of invocations of risk-aversive norms is set to a beta distributed random number with its mean as in the previous sentence (see also C.IV.ii.33.).

IV.ii.3 `NDW` is the weight of the normative drive as opposed to the individual drive, the former being calculated from the saliences of each norm and the resulting propensity to take one of the possible actions, the individual drive being the calculated risk (**0.3**, 0.6, 0.9).

IV.ii.4 `initial-subareas` (13) determines how many subregions (`subareas`) are displayed.

IV.ii.5 The combination of `constant-seed`, `initial-subareas` and `total-population` allows for different topographies.

IV.ii.6 `infected-from-abroad` (`none`) determines whether it is possible that some agents are infected from abroad (as if they had just returned from holidays abroad); in this case about two percent of the population become infected and infectious from the second to the eleventh simulated month (**seasonal**, alternatively: in months 7 to 8 or in months 9 to 11 (`holidays`); such an event is logged in the output area).

IV.ii.7 `superspread?` (false) determines whether it is possible that some agents with a high value of `conviviality * carelessness` convene an event where 50 agents from a small neighbourhood come close together (within the radius of one patch size), thus increasing the probability of being infected just because there are more infectious agents in their close neighbourhood albeit for a short period. Such events are logged in the `output` window together with the number of infections generated at this occasion.

IV.ii.8 `period` shows whether the model is in lockdown mode or in normal mode; this can be switched automatically or by the user.

IV.ii.9 `max-months` (24) determines the maximum duration of the simulation run.

IV.ii.10 `far-commuter` (0.20) determines the ratio of agents which can move up to `long-distance` patches at a time.

IV.ii.11 `short-distance` (2) determines how far (at most) the ordinary agents can walk at a time (in units of patches).

IV.ii.12 `long-distance` (50) determines how far (at most) the far-commuting agents can travel at a time (in units of patches).

IV.ii.13 `auto-lockdown?` (`off`) determines whether the model can enter the lockdown mode automatically. "Lockdown" means that the distances up to which agents may move are lowered to one half of its normal value (`internal`, `external` means that infections from abroad are prevented).

IV.ii.14 `lockdown-at` (50, 100) determines the limit for entering the lockdown: when the number of new infectious agents per 100,000 during the paast seven days (the *incidence rate*, see also C.I.ii.12.) becomes greater than this limit, lockdown starts at the next midnight.

IV.ii.15 `lockdown-end-at` (35, 70 as a percentage of the `lockdown-at` value) determines the limit for exiting the automatic lockdown: when the *incidence rate* becomes less than this limit, lockdown ends at the next midnight.

IV.ii.16 `distance-for-infection` (0.5) is the maximum distance within which infections are possible. This distance is measured as a multiple of the mean distance between an agent and its nearest neighbour, not in patch size (see C.IV.ii.5.).

IV.ii.17 `infections-per-contact` (0.05) is the ratio between successful infections and contacts. In some scenarios, it can change over time and between agents.

IV.ii.18 `mutation-rate` (0.0, **0.05**) is a small positive number which determines the change of the individual variable `virus-infectiosity`. This is initialised in every agent (except `person-zero` and agents which represent people returning infected from abroad) to 0.0, and for an newly infected agent it is set to the `virus-infectiosity` of the respective infector agent and increased (or decreased) by a uniform random variable between `-` `mutation-rate` and `mutation-rate` but never excceds 1.0. This usually leads to a slow increase of the mean and median of the distribution of `virus-infectiosity` of the currently infectious agents.

IV.ii.19 `tested-per-infected` (1.0) is the ratio between tested (and, hence, reportable) infections and all infections.

IV.ii.20 `asymptomatic-per-infected` (0.65) is the ratio between the number of agents which do not show symptoms resulting in being hospitalised and the number of all agents.

IV.ii.21 `case-fatality-rate` (0.04) is the ratio between all deceased and all infected.

IV.ii.22 `time-to-infectiosity-min` (0) and `-max` (4) is the range of a uniform distribution for the time between being infected and being infectious (in days).

IV.ii.23 `incubation-time-min` (2) and `-max` (20) is the range of a uniform distribution for the time between between being infected and showing symptoms of illness that lead to hospitalisation (in days).

IV.ii.24 `illness-duration-min` (12) and `-max` (30) is the range of a uniform distribution for the time between falling ill and recovering or death (in days).

IV.ii.25 `time-till-immunity-min` (0) and `-max` (2) is the range of a uniform distribution for the time between recovery and immunity (in days),

IV.ii.26 `immunity-duration-min` (24) and `-max` (30) is the range of a uniform distribution for the time between recovering and becoming susceptible after loss of immunity (in months). This means that instandard runs there is no loss of immunity at all.

IV.ii.27 `reporting-delay` (0.0) is the upper end of the range of a uniform distribution determining when a test result is reported; over the weekend further delays are programmed (see code). This means that in standard runs there is only a weekend delay.

IV.ii.28 `constant-seed` determines whether during the setup (`setup-constant`) or in both parts (`all-constant`) of the simulation a predefined seed is used or whether each run is controlled by a seed derived from the time when the runs starts.

IV.ii.29 `setup-seed?` provides a standard random number generator seed when `constant-seed` is `setup-constant` or `all-constant` (the seed is then **-1941496835** or the number taken from the respective input widget), otherwise each run starts with a different seed provided by NetLogo (dependent on the current date and time down to the millisecond); `setup-constant` results in a standard topography and distribution of the agents over the topography.

IV.ii.30 `go-seed?` provides a standard random number generator seed when `constant-seed` is `all-constant` (the seed is then **-1637705880** or the number taken from the respective input widget), otherwise after the setup each run starts with a different seed provided by NetLogo (dependent on the current date and time down to the millisecond).

IV.ii.31 `zero-person` determines whether `person-zero` occurs in the `densest`, `sparsest` or a `random subarea`.

IV.ii.32 `interruptible?` (off) determines whether the simulation run can be interrupted at the end of each simulated day. The switch can be changed at any time, and at midnight of the next simulated day a chooser opens which allows the user to determine how the simulation should go on. The options are

- `continue` until next midnight;
- `go on forever`: continue forever;

- `print profile and continue forever`: (only for debugging purposes) print a profile (duration of all procedures) and continue;
- `halt and print profile and result!`: end the simulation run and output all collected information;
- `spread`: call the procedure `move-an-infectious-far-away`, i.e. select an infectious agent randomly and put it on a random patch;
- `normal time`: exit from a user-defined lockdown;
- `lockdown`: start a user-defined lockdown;
- `inspect` a randomly selected infectious agent.

IV.ii.33 `heterogeneity` determines whether personal risk aversion (`conviviality * carelessness`) and/or normative background (`BACKGROUND` and `individual-LNP`)

- are equal for all agents (alternative `none`) or
- are beta distributed individual variables (alternative `wrt risk` between 0 and 1 for both `conviviality` and `carelessness`, alternative `wrt norm` between `BACKGROUND - 10` and `BACKGROUND + 10`);
- for the alternative `both` (the standard) all three instance variables are random.

IV.ii.34 `vaccination?` determines whether after some time (`vaccination-start`, standard 355) the first agents will receive a vaccination; for every simulated day a number of doses is made available. How many these are (`doses-per-day`) depends on the parameter `vacc-lasts-days`, the number of days until the total population could be vaccinated. Hence, `doses-per-day` is the `total-population` divided by `vacc-lasts-days`. As only susceptible agents willing to be vaccinated (`carelessness < 0.7`) receive doses, remaining doses are used on the next day. The numbers of administered and refused vaccinations are reported, the number of doses available on each day is only reported to a monitor.

IV.iii. *How were submodels designed or chosen, and how were they parameterised and then tested?* The calendar submodel just organises a discrete-event model with the help of the `time` extension. The events scheduled by the `people` agents generalise a SEIRS model.

References

Adam, D.C., Wu, P., Wong, J.Y., Lau, E.H.Y., Tsang, T.K., Cauchemez, S., ... Cowling, B.J. (2020, sep). Clustering and superspreading potential of SARS-CoV-2 infections in hong kong. *Nature Medicine*. 10.1038/s41591-020-1092-0

Agrawal, M., Kanitkar, M., Vidyasagar, M. (2021). *Sutra: A novel approach to modelling pandemics with asymptomatic and undetected patients, and applications to covid-19.*

an der Heiden, M., & Hamouda, O. (2020). Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland — Nowcasting. *Epidemiologisches Bulletin*(17), 10–16. 10.25646/6692.4

Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., Villatoro, D. (2013). Punish and voice: Punishment enhances cooperation when combined with norm-signalling. *PLoS ONE*, *8*(6), e64941. 10.1371/journal.pone.0064941

Andrighetto, G., & Castelfranchi, C. (2013). Norm compliance: the prescriptive power of normative actions. *Paradigmi*, *[6]*(2), 120–135. 10.3280/PARA2013-002011

Blumberg, S., Funk, S., Pulliam, J.R.C. (2014). Detecting differential transmissibilities that affect the size of self-limited outbreaks. *PLOS Pathogens*, *10*(10), e1004452. 10.1371/journal.ppat.1004452

Bundesamt für Gesundheit (2021, August). *Weekly situation report on the epidemiological situation in Switzerland and the Principality of Liechtenstein.* Retrieved from https://www.bag.admin.ch/dam/bag/en/dokumente/mt/k-und-i/aktuelle-ausbrueche-pandemien/2019-nCoV/covid-19-basisdaten-fallzahlen.csv.download.csv/Data%20on%20laboratory%20findings%20and%20deaths.csv (in German, last download 2021-08-26)

Bundesministerium für Soziales, P.u.K., Gesundheit (2021). *Katalog covid-19: Zeitliche darstellung von daten zu covid19-fällen je bezirk.* Retrieved from https://www.data.gv.at/katalog/dataset/covid-19-zeitliche-darstellung-von-daten-zu-covid19-fallen-je-bezirk/resource/9eb08d45-ff99-40f1-90cd-7b3659b0bc8d (last download 2021-08-26)

Cialdini, R.B., Reno, R.R., Kallgren, C.A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public

places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. 10.1037/0022-3514.58.6.1015

Cochran, J.J. (2020, June). Why we need more coronavirus tests than we think we need. *Significance (Oxford, England)*, *17*(3), 14—15. Retrieved from https://europepmc.org/articles/PMC7283702 10.1111/1740-9713.01398

Conte, R., Andrighetto, G., & Campennì, M. (Eds.). (2013). *Minding norms. mechanisms and dynamics of social order in agent societies.* Oxford: Oxford University Press.

Elsenbroich, C., Anzola, D., & Gilbert, N. (Eds.). (2016). *Social dimensions of organised crime. modelling the dynamics of extortion rackets.* Cham: Springer. 10.1007/978-3-319-45169-5

Endo, A., Abbott, S., Kucharski, A.J., Funk, S. (2020). Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome Open Research*. 10.12688/wellcomeopenres.15842.3

Epstein, J.M., & Axtell, R. (1996). *Growing artificial societies – social science from the bottom up.* Cambridge, MA: MIT Press.

Figge, M.T. (2005, May). Stochastic discrete event simulation of germinal center reactions. *Phys. Rev. E*, *71*(5), 051907. 10.1103/PhysRevE.71.051907

Gaudou, B., Huynh, N.Q., Philippon, D., Brugière, A., Chapuis, K., Taillandier, P., ... Drogoul, A. (2020, sep). COMOKIT: A modeling kit to understand, analyze, and compare the impacts of mitigation policies against the COVID-19 epidemic at the scale of a city. *Frontiers in Public Health*, *8*, 587. 10.3389/fpubh.2020.563247

Gelfand, M.J., Jackson, J.C., Pan, X., Nau, D., Pieper, D., Denison, E., ... Wang, M. (2021, jan). The relationship between cultural tightness–looseness and COVID-19 cases and deaths: a global analysis. *The Lancet Planetary Health*. 10.1016/S2542-5196(20)30301-6

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., ... DeAngelis, D.L. (2006, sep). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, *198*(1-2), 115–126. 10.1016/j.ecolmodel.2006.04.023

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F. (2010, nov). The ODD protocol: a review and first update. *Ecological Modelling*, *221*(23), 2760-2768.
10.1016/j.ecolmodel.2010.08.019

Hoebel, J., Busch, M.A., Grabka, M.M., Zinn, S., Allen, J., Gößwald, A., ... Lampert, T. (2021). *Seroepidemiologische Studie zur bundesweiten Verbreitung von SARS-CoV-2 in Deutschland — CORONA-MONITORING bundesweit (RKI-SOEP-Studie)* (Tech. Rep.). 10.25646/7852

Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Rico, M.S., ... Leleu, H. (2020, July). A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine*, *26*(9), 1417–1421.
10.1038/s41591-020-1001-6

Hoertel, N., Blachier, M., Sánchez-Rico, M., Limosin, F., Leleu, H. (2021, jan). Impact of the timing and adherence to face mask use on the course of the COVID-19 epidemic in france. *Journal of Travel Medicine*, *28*(2).
10.1093/jtm/taab016

Hsieh, Y.-H., Chang, H.-L., Lee, J.-Y. (2004, June). SARS epidemiology modeling. *Emerging Infectious Diseases*, *10*. Retrieved from http://dx.doi.org/10.3201/eid1006.031023
10.3201/eid1006.031023

Istituto Nazionale di Statistica (2020, August). *Primi risultati dell'indagine di sieroprevalenza sul sars-cov-2.* Retrieved from https://www.istat.it/it/files//2020/08/ReportPrimiRisultatiIndagineSiero.pdf (last downloaded on 2020-10-02)

Lauring, A.S., & Hodcroft, E.B. (2021, January). Genetic variants of sars-cov-2—what do they mean? *JAMA*.
10.1001/jama.2020.27124001/jama.2020.27124

Lloyd-Smith, J.O. (2007, 02). Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLOS ONE*, *2*(2), 1-8. Retrieved from https://doi.org/10.1371/journal.pone.0000180
10.1371/journal.pone.0000180

Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., Getz, W.M. (2005, nov). Superspreading and the effect of individualvariation on disease emergence. *Nature*, *438*(7066), 355–359.
10.1038/nature04153

Moore, K.A., Lipsitch, M., Barry, J.M., Osterholm, M.T. (2020, April). *COVID-19: The CIDRAP viewpoint. part 1: The future of*

the COVID-19 pandemic: Lessons learned from pandemic influenza. Retrieved from https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-covid19-viewpoint-part1_0.pdf. (last downloaded 2021-01-14)

Müller, B., Bohn, F., Dressler, G., Groeneveld, J., Klassert, C., Matrtin, R., ... Schwarz, N. (2013). Describing human decisions in agent-based models — ODD+D, an extension of the ODD protocol. *Environmental Modelling & Software*, *48*, 37–48.
10.1016/j.envsoft.2013.06.003

Nistal, R., de la Sen, M., Gabirondo, J., Alonso-Quesada, S., Garrido, A.J., Garrido, I. (2021, jul). A study on COVID-19 incidence in Europe through two SEIR epidemic models which consider mixed contagions from asymptomatic and symptomatic individuals. *Applied Sciences*, *11*(14), 6266.
10.3390/app11146266

Perumalla, K.S., & Seal, S.K. (2012, jul). Discrete event modeling and massively parallel execution of epidemic outbreak phenomena. *Simulation*, *88*(7), 768–783.
10.1177/0037549711413001

Pescarmona, G., Terna, P., Acquadro, A., Pescarmona, P., Russo, G., Sulis, E., Terna, S. (2021, August). *An agent-based model of covid-19 diffusion to plan and evaluate intervention policies.* Retrieved from https://arxiv.org/abs/2108.08885

Pollán, M., Pérez-Gómez, B., Pastor-Barriuso, R., Oteo, J., Hernán, M.A., Pérez-Olmeda, M., ... Yotti, R. (2020). Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet*, *396*, 535–544.
10.1016/S0140-6736(20)31483-5

Robert-Koch-Institut (2021, Februar). *CSV mit den aktuellen Covid-19 Infektionen pro Tag (Zeitreihe).* Berlin. Retrieved from https://opendata.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0.csv (Data licence Germany — attribution — Version 2.0 www.govdata.de/dl-de/by-2-0; data retrieved 2021-08-26)

Squazzoni, F., Polhill, J.G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G., ... Gilbert, N. (2020). Computational models that matter during a global pandemic outbreak: A call to action. *Journal of Artificial Societies and Social Simulation*, *23*(2).
10.18564/jasss.4298

Statistisches Bundesamt (2020, September). *Daten aus dem Gemeindeverzeichnis. Kreisfreie Städte und Landkreise nach Fläche, Bevölkerung und Bevölkerungsdichte. Gebietsstand: 31.12.2019.* Retrieved from https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/04-kreise.html (retrieved 2020-09-18)

Steinmann, P., Wang, J.R., van Voorn, G.A., Kwakkel, J.H. (2020). Don't try to predict COVID-19. if you must, use Deep Uncertainty methods. *Review of Artificial Societies and Social Simulation*. Retrieved from https://rofasss.org/2020/04/17/deep-uncertainty/ (17th April 2020)

Terna, P., Pescarmona, G., Acquadro, A., Pescarmona, P., Russo, G., Terna., S. (2020). *An agent-based model of the diffusione of covid-19 using netlogo.* Retrieved from https://terna.to.it/simul/SIsaR.html (Last update Sep 14th, 2020)

Troitzsch, K.G. (2018). Can lawlike rules emerge without the intervention of legislators? *Frontiers in Sociology*, *3*, 2.
10.3389/fsoc.2018.00002

Vermeulen, B., Pyka, A., Müller, M., Kugelmeier, F.U. (2020, August). *An agent-based policy laboratory for COVID-19 containment strategies.* Retrieved from http://sowi.st-ursula-attendorn.de/politiklabor_a.htm

Wallinga, J., & Lipsitch, M. (2007, February). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B*, *274*, 599–604.
doi:10.1098/rspb.2006.3754

Wu, S.L., Mertens, A.N., Crider, Y.S., Nguyen, A., Pokpongkiat, N.N., Djajadi, S., ... Benjamin-Chung, J. (2020, September). Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications*, *11*, 4507. Retrieved from https://doi.org/10.1038/s41467-020-18272-4
10.1038/s41467-020-18272-4

Young, B.E., Fong, S.-W., Chan, Y.-H., Mak, T.-M., Ang, L.W., Anderson, D.E., ... Ng, L.F.P. (2020, August). Effects of a major deletion in the sars-cov-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet*, *396*, 603–611.
10.1016/S0140-6736(20)31757-8

Zhou, G., & Yan, G. (2003, December). Severe acute respiratory syndrome epidemic in asia. *Emerging Infectious Diseaeses*. (last retrieved on 2020-10-21)
10.3201/eid0912.030382

Klaus G. Troitzsch studied sociology and political science in Cologne and Hamburg. After having taken his PhD in political science from the University of Hamburg, he first worked in an election research project and from 1986 as full professor of computer applications in the social sciences at the University of Koblenz-Landau. In March 2012 he retired but continues his academic activities.

He is author, co-author, and co-editor of a number of books on simulation, author of numerous articles in social simulation, and he organised or co-organised a number of national and international conferences in social simulation.