2021, 7 (1)

ARGUMENTA

The Journal of the Italian Society for Analytic Philosophy

First published 2021 by the University of Sassari

Produced and designed for digital publication by the *Argumenta* Staff

Argumenta is the official journal of the Italian Society for Analytic Philosophy (SIFA). It was founded in 2014 in response to a common demand for the creation of an Italian journal explicitly devoted to the publication of high quality research in analytic philosophy. From the beginning *Argumenta* was conceived as an international journal, and has benefitted from the cooperation of some of the most distinguished Italian and non-Italian scholars in all areas of analytic philosophy.

Contents

Editorial

If we were seeking consolation for the huge difficulties that have plagued us since the COVID-19 pandemic broke out, we might draw some comfort from the breadth and depth of the discussions that it has given rise to, for all that these have often led to the expression and propagation of misleading views. The present number of *Argumenta* opens with a Special Issue that intends to take stock of the tenor and quality of these discussions, bringing to light the awareness that we have been acquiring over the last few months about our overall cognitive situation.

The Special Issue, edited by Margherita Benzi, Raffaella Campaner and Francesco Barone-Adesi, is entitled *Modelling the Covid-19 Pandemic: Epidemiological, Epistemological, and Ethical Challenges*, and draws attention to the intricacy of formulating models that take into account the many factors at work in a pandemic. The result is an up-to-date discussion of these aspects that, by focusing on epistemological and epidemiological views, seek to identify what epistemic virtues should guide the choice of models for explanatory and predictive purposes.

The present number also includes three articles that have already appeared in 'early view' (by Fabio Bacchini and Nicola Piras, John Biro, and Anna Ichino and Juha Räikkä), and that have already made and will continue to make significant contributions to discussion in their respective fields.

Argumenta 7, 1 (2021): 3-4
ISSN 2465-2334

The number is then topped off by the section of Book Reviews. We are proud to offer readers three new thoughtful reviews of as many interesting books.

Finally, I would like to thank all the colleagues who have acted as external referees, the Assistant Editors, the Editor of the Book Reviews, the members of the Editorial Board, and the Editors of the Special Issue. All of them have been very generous with their work, advice, and suggestions.

As usual, the articles appearing in *Argumenta* are freely accessible and freely downloadable, therefore it only remains to wish you:

Buona lettura!

Massimo Dell'Utri
Editor-in-Chief

Argumenta 7, 1 (2021)
Special Issue

Modelling the COVID-19 Pandemic:

Epidemiological, Epistemological, and Ethical Challenges

Edited by

Margherita Benzi, Francesco Barone-Adesi,
Raffaella Campaner

The Journal of the Italian Society for Analytic Philosophy

Contents

Introduction:
COVID-19 Models and the Difficult Balance between Methods and Values

Margherita Benzi
University of Piemonte Orientale

Francesco Barone-Adesi
University of Piemonte Orientale
CRIMEDIM—Research Center in Emergency and Disaster Medicine

Raffaella Campaner
University of Bologna

The COVID-19 pandemic had an unprecedented impact not only on the socio-economic and political conditions worldwide but also on the practices of the scientific community and on the public image of science itself. The scientific community suddenly found itself in the spotlight and was pressured to rapidly produce evidence applicable to the management of the present health crisis. This in turn had some unexpected consequences, among which an increase of the publication speed and sometimes a decrease of the quality of peer review (see e.g., Chan 2020). At the same time, the public discussion of scientific issues related to COVID-19 among an audience often lacking the appropriate knowledge of the characteristics of modern science (e.g., critical reasoning, hypothetical nature of research, the role of uncertainty, …), was associated with the emergence of extreme stances in the population. These include distrust and refusal of the scientific authority, on the one side, and acritical scientism, on the other. It is plausible that such attitudes may have affected in a negative way the behaviors of people and their compliance to the preventive measures put in place to tackle the pandemic. In this respect, diffusion of knowledge about the way science really works and an increase of active participation in the critical-methodological discussion could be important for better managing the pandemic in the future. Who might start fostering a constructive and fruitful dialogue, and how to do so, is one of the crucial concerns from which this issue originates. Indeed, the pandemic has promoted an intense discussion between epidemiologists and philosophers of science (mainly, but not surprisingly, philosophers of medicine and epidemiology). This debate was mostly focused on methodological issues. The COVID-19 pandemic has obviously also prompted reflections by other types of scientists and philosophers. A notable example of plurality of perspectives is Boniolo and Onaga (2021). In this special issue we decided to focus on the epistemological and epidemiological views because we think they have proved of

Argumenta 7, 1 (2021): 9-17
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.int

central importance in the last couple of years. Moreover, they can be of help—and often they cannot be ignored—also in evaluating contributions coming from other fields, such as bioethics, political philosophy, health policy assessment, and communication theory.

Since the very start of the pandemic, the lion's share of the debate was on the use and the utility of epidemiological models for the prediction of pandemic evolution and for supporting the decisions regarding the introduction of public health measures such as contact tracing, quarantine, and lockdowns. A thorough reflection on these models requires specifying what is the phenomenon to be modelled, the variables and the (causal) relations to consider, the optimal degree of realism/idealization, detail/abstraction of the model necessary to provide useful predictions and effective control strategies, the methods to evaluate the performance of models, the best way to communicate the results to inform policy decisions.

Epidemiologists use very different types of models to answer research questions typical of their field, but the most used ones are arguably the regression models. These methods are largely empiric, in the sense that they usually do not rely on strong a priori assumptions regarding the theoretical mechanisms behind the phenomenon being studied, but rather evaluate the association between independent variables (e.g., age, gender, smoking habit) and one or more dependent variables (e.g., risk of death) through a black-box, theory-free, approach. Some famous examples are the models for cardiovascular risk derived by the Framingham study (Mahmood 2014) and the plethora of models of cancer (Peto 2001), which are probably some of the hallmarks of modern epidemiology (Galea 2010). This tradition is rooted in what is sometimes defined as the "etiologic epidemiology of non-communicable diseases". During the Fifties, a new causal paradigm to explain the relationship between smoking and lung cancer was proposed and was then extended to most non-communicable diseases. This paradigm lies on the concept of risk factor to define a not necessary and not sufficient cause that increases the probability of an event to occur and found its consolidation in the so-called Austin Bradford Hill criteria, which base the evaluation of causality on observational data (Hill 1965). This second tradition of epidemiologic thinking evolved through the years, developing increasingly sophisticated methodological approaches (e.g., DAG, counterfactuals, etc.) to try to overcome the "original sin of non-randomization" and provide more robust causal inference from observational data (Vanderbroucke et al. 2016).

Interestingly, one of the effects of the COVID-19 pandemic in the scientific-philosophical debate has been to put in the spotlight a different type of models, which was substantially less common in epidemiology, namely the mathematical models of infectious disease. While the first examples of these models date back to the beginning of the 20th century, it was from the Seventies onward that these models gained a central role in infectious disease epidemiology (Koopman 2015). The main feature of these models is that they allow us to take into account the complex transmission dynamics of infective agents among the population, which is impossible using normal regression models.

Loosely speaking, models used for the prediction of COVID-19 trends can be divided into three broad groups (Adams 2020): compartmental models (e.g., SEIR models), individually-oriented models (e.g., Agent Based Models), and curve-fitting approaches. The first two groups include mathematical models that simulate the behavior of an epidemic based on a priori set of parameters' values.

The last group is more heterogeneous and includes models that estimate the values for the parameters directly from observed data. Differently from the models of the first two groups, curve fitting models are usually empiric (e.g., regression models based on the logistic function) or have a degree of theorization regarding the diffusion mechanisms of the agent substantially lower compared to compartmental models and agent-based models. This classification is obviously an oversimplification. In the real world, some models have features of both compartmental and individual models, and curve-fitting sometimes is carried out using compartmental models.

From the philosophical point of view, the debate on the use of models during the COVID-19 pandemic represents an interesting case study for at least three main reasons:

1) At the beginning of the pandemic, basic knowledge on SARS-Cov2 and COVID-19 (e.g., transmission rate, mortality rate, routes of transmissions, number of asymptomatic subjects in the population, and their role in the spread of the disease) was largely lacking (Bellan et al. 2020, Yanes-Lane et al. 2020, Caristia et al. 2020).
2) It was necessary to rapidly decide whether to implement public health interventions (i.e., lockdown) that would have substantially reduced personal freedom and possibly also had negative socio-economic consequences in the population.
3) The only quantitative results on which basing policy decisions were derived by complex mathematical models, lying on several assumptions, whose reliability was somewhat dubious even among the scientific community.

From this point of view, it is interesting to go back to a discussion on the reliability of mathematical models in COVID-19 started by the philosopher of medicine Jonathan Fuller during the first wave of the pandemic in May 2020. In a series of articles published in the *Boston Review*, Fuller (2020a, 2020b) suggests that two different traditions of epidemiological thinking, namely clinical epidemiology and public health epidemiology, have very different stances regarding what methodological approaches are to be considered acceptable to inform public health decisions during the pandemic. The former mainly refers to the principles of the movement known as Evidence-Based Medicine or EBM and the latter overlaps to what we previously defined as etiologic epidemiology. In particular, Fuller referred to John Ioannidis, professor of epidemiology at Stanford and well-known for his provocative meta-scientific contributions to the discipline, and Marc Lipsitch, professor of epidemiology at Harvard, as "champions" of the two traditions (Ioannidis 2020a, 2020b, Lipsitch 2020). The position of Ioannidis, which would be then largely stigmatized, was critical toward the use of models to support decisions on how to manage the pandemic, as they were felt by the author as based on low-quality data and based on types of studies not meeting the standards required by EBM. On the opposite, Lipsitch noted that in situations where uncertainty is high, time is scarce and stakes are high, it is necessary to consider any type of knowledge that could be useful to generate hypotheses and make predictions, including the theoretical knowledge coming from fields different from epidemiology and in general "weak" forms of evidence.

After almost two years from the beginning of this debate, the apparent contraposition from these two positions somewhat faded away (Fuller 2020b). It has become much clearer that it is useless to stick to dogmatic views about what constitutes evidence, and that, on the contrary, the peculiarity of the present situation requires exploring novel ways to better understand such a complex phenomenon and consequently to envision possible effective interventions. However, a general epistemic question remains somewhat unanswered: is there any way to thoroughly evaluate the knowledge coming from complex models, which are full of untestable (and often implicit) assumptions and approximations, and use it to inform public health decisions? This question, in turn, calls for further reflections regarding the values involved in such decisions ("is an intervention doing the best for whom?"). On this topic, see the conclusive remarks in Fuller (2020b, 2021) and the different communication strategies (e.g., is intelligibility for decision-makers a virtue of a model?). If anything, papers included in this special issue witness how the COVID-19 pandemic has discouraged a value-free vision of models and whole science in general.

The discussion launched by Fuller was not the only philosophical debate prompted by the pandemic. The journal *Nature* (June 2020) published a *Manifesto* for the correct use of models, written by a group of scientists and philosophers. The authors stressed that in many cases, the epistemic and social aspects of the use of modelling and of using models are not fully distinguishable. Model users should keep in mind that no one model can serve all purposes, as "results from the models will at least partly reflect the interests, disciplinary orientations, and biases of the developers" (Saltelli et al. 2020). Moreover, models can be inspired by different sets of values that should be explicitly declared by the modellers. In their conclusion*,* the authors made a plea for two things: (1) using models to question the world, rather than to provide definitive answers, and (2) allowing broad participation in the formulation and reflection on models. These two themes appear prominently in the contributions to this special issue, alongside "classical" themes in the philosophy of science. Consequently, the remarks on models that we present here can contribute to shedding light on how the pandemic has affected the way of doing science and philosophy of science.

The first of the aspects we have mentioned raises the question of the type of models adopted in the study of the spread of COVID-19. Olaf Damman's paper analyzes the explanatory function of a particular type of simulative model adopted in the pandemic, the ABM (Agent-Based Models). ABMs are non-deterministic models that simulate changes in populations over time based on the behavior of individual agents who interact according to rules defined in the program. Damman discusses three philosophical aspects of ABMs: their usefulness for causal inference as models of causal mechanisms, the question of whether they represent truly emergent phenomena, and their explanatory role. With regard to the third point, Damman argues that ABMs provide a particular kind of explanation, etio-prognostic explanation, of illness occurrence and outcome.

Till Grüne-Yanoff presents a reflection on what epistemic virtues should guide the choice of models, referring to a case study, the choice of a compartmental model by the Public Health Agency of Sweden (Folkhälsomyndigheten, or FoHM) in the first part of the pandemic. Grüne-Yanoff analyzes the considerations justifying the choice of a compartmental model, instead of an ABM model, by FoHM modellers. Although ABM can guarantee a higher degree of similarity to the target, compartmental models are simpler. The author argues in favor of the trade-off between similarity and simplicity discussing several epistemic virtues related to the latter. Interestingly, he includes ease of communication among epistemic virtues, which seems strange, since in general ease of communication does not seem to concern the creation of knowledge and therefore does not constitute an epistemic value. However, argues Grüne-Yanoff, ease of communication becomes an epistemic value when one considers the broad interdisciplinary nature of the teams of those called upon to build models to counter the pandemic.

As known, models can pursue various goals. Among them, representing causal relations is undoubtedly one of the main targets of models that are meant to drive decisions in the struggle against the pandemic: were we aware of the genuinely causal relations bringing about the disease, we would be able to intervene to either prevent or, at least, treat and cure it. However, looking for causes is not per se an easy task, nor does it rely on any univocal and universally shared understanding of what causes ultimately are and, even more so, where and how they are to be sought. Federico Boem draws some epistemologically relevant differences between proximate and ultimate causes, where the former can appear more clearly in front of us at present, whereas the latter are to be understood from an ecological, evolutionary, and socio-economic standpoint. His contribution advocates the idea that in such contexts as the COVID-19 pandemic modelling needs to combine different sorts of causes, including evolutionary and socio-economic factors, to reach an integrated understanding. Daniel Auker-Howlett and Jon Williamson, on their hand, focus their reflections on vaccination against COVID-19, stressing how local and social mechanisms can make a difference with respect to the assessment and refinement of vaccination intake interventions. Starting from recent epistemological reflections on causal evidence and what it can amount to in the context of Evidence Based Medicine—and, more specifically, stressing the advantages of the approach known as EMB+—Auker-Howlett and Williamson point out how the gathering of mechanistic knowledge and the elaboration of detailed mechanistic models can offer benefits for research on vaccination and lead to more effective interventions. Considerations on the relevance of genuinely mechanistic knowledge of how COVID-19 actually behaves are hence inserted in the wider debate on EBM, its pros and limits, and stress the importance of going beyond correlational knowledge.

Whereas Auker-Howlett and Williamson's paper highlight how, in the end, the applicability of results can be an extremely relevant guiding principle in the scientific enterprise, the contribution by Annibale Biggeri and Andrea Saltelli questions another epistemic virtue, precision, and in particular that expressed by the descriptions and predictions of the pandemic. However, the two authors point out that numbers are based on various assumptions, of which they do not guarantee the absence of bias. An emblematic example of how the precision of the numbers can mask controversial assumptions is given by the calculation of excess mortality during the first wave of the pandemic, defined as the difference between the total number of deaths and the expected number of deaths—i.e., the counterfactual number of deaths it would have been observed in absence of pandemic. As this indicator "depends strongly on the calculation of the expected death counts", it is strongly dependent on the assumptions on which the model is based. Here, the authors present a case study—a set of different estimates of the excess mortality during the first wave of the pandemic in Italy—to show how they varied under different methodological choices. In the conclusions, they suggest that the stimulus for a more careful analysis of the assumptions which underlie models could come from a new approach to the relationship between science and society.

These issues are also the focus of the contribution of Paolo Vineis et al. The authors highlight how in the description of epidemics idiographic, circumstantial aspects, such as chance, historical and geographical context, ..., count at least as much as nomothetic ones. Among the characteristic aspects of a pandemic there are factors belonging to heterogeneous categories but often linked by a relationship of mutual influence. The authors refer to Pierre Bourdieu's categories of different kinds of capitals: economic, social, and cultural capitals, to which should be added a "biological capital", including an "immunological capital". They point out that these categories should be considered in the construction of models, highlighting the ethical and political burden of models and measures aimed at countering the pandemic. How to make explicit and possibly formalize the consideration of values in model building is, the authors suggest, one of the new tasks that the pandemic seems to be assigning to us.

The contribution of Virginia Ghiara discusses even with more detail the particular aspects of the pandemic concerning vaccination policies. Ghiara emphasizes how the consideration of the 'mechanisms' defended by the authors who recognize themselves in the strand of research known as EBM+ is suitable for the assessment of effectiveness and efficiency of vaccines—both in the evaluation of potential pathways and future directions of research and in the analysis of vaccination behaviors, fundamental to design vaccination campaigns. A correct evaluation of these behaviors requires an analysis of the mechanisms of facilitation and impediment that influence vaccination behaviors in different social and geographical contexts. In this regard, Ghiara illustrates how the World Health Organization is promoting the collection of mechanistic evidence to understand the potential efficacy of particular vaccination interventions in different contexts.

Elena Rocca and Birgitta Grundmark devote their attention to pharmacovigilance, i.e., "the science of detecting and assessing possible adverse reactions from medical interventions". They discuss how the peculiar features of the COVID-19 pandemic—which, on the one hand, has provided us with unprecedented amounts of data and, on the other hand, has forced us to struggle with varied and uncertain evidence—call for a deeper reflection on the need of contributions from epistemology, ethics and philosophy of science in the understanding and managing of a crisis. Using critical thinking to tackle evidence and scientific success, one can better cope with uncertainty and deal with its challenges.

Given that, as recalled above, COVID-19 models have not only provided a theoretical understanding of the disease, but also guided political and economic decisions worldwide, and more or less successfully so, the very ideas of expertise and trust in science need to be brought into focus. Such ideas can constitute an essential terrain to discuss the interplay of epistemic and non-epistemic values in the construction and communication of scientific knowledge, bringing to light, on the one hand, the constraints under which science is pursued and the limits that can derive from them, and, on the other hand, why it keeps on being the most reliable form of knowledge. Why should society trust experts, and which experts should it trust? How is expertise achieved, how is it assessed, and what role does it play in the understanding of the pandemic and in the dissemination of scientific knowledge on the disease? Carlo Martini addresses questions along these lines, investigating possible ways of interaction between a model and expert options and different directions in which expert judgments can impact choices and uses of models themselves.

Cecilia Nardini and Fridolin Gross approach the topic of shared science from another perspective, that of bottom-up initiatives of independent citizens engaged in data production, data review, and, to some extent, model production. Such a perspective is typical of the so-called 'citizen science', to which current literature attributes two alternative modalities: the direct contribution of citizens to data collection under scientists' guidance, and the pressure on the scientific community to raise awareness on socio-political issues. Nardini and Gross analyze the activities of the community of non-professional users of COVID-related data on the software sharing platform GitHub and show that they cannot be framed in the two recognized strands of citizen science. Instead, they seem motivated by individual curiosity and the intent to improve the information received from the media.

Science belongs to society as a whole, and hence to citizens and groups, which have expectations with respect to science and its impact on their lives. Nicolò Gaj and Giuseppe Lodico's contribution deals with the dissemination and popularization of scientific outcomes regarding COVID-19, discussing scientism—as "a stance identifying science as the only reliable source of legitimate knowledge"—and its relations with naturalism and the debate on science's unity/disunity. A deeper analysis of such relations, with a stress on the plurality of methods and concepts adopted by science, can foster a better understanding of science's actual inner working and, hence, a more balanced public outlook on science, what it is and what it is not, its credibility in the social scenario, what we can and cannot expect from it. A subtle analysis of the whole range of methods and concepts put in place also in tackling the pandemic brings with it also a better understanding of what data and evidence amount to, how they can be gathered and amalgamated.

Alongside the issues of how data are produced and transmitted, we find the problem of how data are received. Years of research in the behavioral and cognitive sciences have made us aware that our perception of information and its use in decision-making deviate from the canons of classical rationality. Among the consequences of the COVID-19 pandemic, we can count a remarkable commitment of the social sciences in shaping the behavior of citizens towards the measures adopted by public health to ensure cognitive architectures capable of promoting bias-free behaviors. Not surprisingly, anti-vaccine behaviors have been a key component of this type of research. Stefano Calboli and Vincenzo Fano's contribution fits within this framework. After presenting what they believe to be the relevant psychological mechanisms in determining vaccine choice, the authors question the effectiveness of policy measures based on economic disincentives to vaccine refusal. The original explanation put forward by the two authors on the instances of the ineffectiveness of such measures is based on the tendency to keep as many options open as possible. The two authors outline an experiment to test their hypothesis, although they conclude with a call for epistemic caution in translating research findings.

References

Adams, J. 2020, "What Are COVID-19 Models Modeling?", *The Society Pages*, April 8, https://thesocietypages.org/specials/what-are-covid-19-models-modeling/ (Accessed September 20, 2021).

Bellan, M., Patti, G., Hayden, E., Azzolina, D., Pirisi, M., Acquaviva, A., Aimaretti, G., Aluffi Valletti, P., Angilletta, R., Arioli, R., Avanzi, G.C., Avino, G., Balbo, P.E., Baldon, G., Baorda, F., Barbero, E., Baricich, A., Barini, M., Barone-Adesi, F., and Battistini, S. 2020, "Fatality Rate and Predictors of Mortality in an Italian Cohort of Hospitalized COVID-19 Patients", *Sci Rep.*, Nov 26;10(1):20731, doi: 10.1038/s41598-020-77698-4.

Boniolo, G., and Onaga, L. (eds.), 2021, Topical collection "Seeing clearly through COVID-19", *History and Philosophy of the Life Sciences*, 43.

Caristia, S., Ferranti, M., Skrami, E., Raffetti, E., Pierannunzio, D., Palladino, R., Carle, F., Saracci, R., Badaloni, C., Barone-Adesi, F., Belleudi, V., and Ancona, C. 2020, "AIE Working Group on the Evaluation of the Effectiveness of Lockdowns. Effect of National and Local Lockdowns on the Control of COVID-19 Pandemic: A Rapid Review", *Epidemiol Prev.*, Sep-Dec; 44(5-6 Suppl 2), 60-68.

Chan, J., Oo, S., Chor, CY.T., Yim, D., Chan, J.S.K., and Harky, A. 2020, "COVID-19 and Literature Evidence: Should We Publish Anything and Everything?", *Acta Biomed.* Sep, 7;91(3):e2020020.

Fuller, J. 2020a, "Models vs. Evidence", *Boston Review*, 1, May, https://bostonreview.net/articles/jonathan-fuller-models-v-evidence/

Fuller, J. 2020b, "From Pandemics Facts to Pandemic Policies", *Boston Review*, 2, June, https://bostonreview.net/articles/jonathan-fuller-poli/

Fuller, J. 2021, "What Are the COVID-19 Models Modeling (Philosophically Speaking)?", *History and Philosophy of the Life Sciences*, 43, 2, 1-5.

Galea, S., Riddle, M., and Kaplan, G.A. 2010, "Causal Thinking and Complex System Approaches in Epidemiology", *Int. J Epidemiol.*, 39, 1, 97-106, doi: 10.1093/ije/dyp296. Epub 2009 Oct 9.

Hill, A.B. 1965, "The Environment and Disease: Association or Causation?", *Proc R Soc Med.*, May 58(5), 295-300.

Ioannidis, J.P. 2020a, "A Fiasco in the Making? As the Coronavirus Pandemic Takes Hold, We Are Making Decisions without Reliable Data", Stat, 17, https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/

Ioannidis, J.P. 2020b, "The Totality of the Evidence", *Boston Review*, 24 May, https://bostonreview.net/articles/john-p-ioannidis-totality-evidence/

Koopman, J.S. 2005, "Infection Transmission Science and Models", *Jpn J Infect Dis.*, Dec; 58(6): S3-8.

Lipsitch, M. 2020, "Good Science Is Good Science", *Boston Review*, 12 May, https://bostonreview.net/articles/marc-lipsitch-good-science-good-science/

Mahmood, S.S., Levy, D., Vasan, R.S., and Wang, T.J. 2014, "The Framingham Heart Study and the Epidemiology of Cardiovascular Disease: A Historical Perspective", *Lancet*, Mar 15;383(9921):999-1008, doi: 10.1016/S0140-6736(13)617 52-3. Epub 2013 Sep 29.

Peto, J. 2001, "Cancer Epidemiology in the Last Century and the Next Decade", *Nature*, May 17; 411(6835), 390-95, doi: 10.1038/35077256.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke, R.Jr., Portaluri, T., Porter, T.M., Puy, A., Rafols, I., Ravetz, J.R., Reinert, E.S., Sarewitz, D., Stark, P.B., Stirling, A., van der Sluijs, J.P., and Vineis, P. 2020, "Five Ways To Ensure That Models Serve Society: A Manifesto", *Nature* 582, 482-84, doi: 10.1038/d41586-020-01812-9.

Vandenbroucke, J.P., Broadbent, A., and Pearce, N. 2016, "Causality and Causal Inference in Epidemiology: The Need for a Pluralistic Approach", *Int J Epidemiol*, Dec1;45(6), 1776-786.

Yanes-Lane, M., Winters, N., Fregonese, F., Bastos, M., Perlman-Arrow, S., Campbell, J.R., and Menzies, D. 2020, "Proportion of Asymptomatic Infection Among COVID-19 Positive Persons and Their Transmission Potential: A Systematic Review and Meta-Analysis", *PLoS One*, Nov 3;15(11): e0241536, doi: 10.1371/journal.pone.0241536.

Agent-Based Models as Etio-Prognostic Explanations

Olaf Dammann

Tufts University

Abstract

Agent-based models (ABMs) are one type of simulation model used in the context of the COVID-19 pandemic. In contrast to equation-based models, ABMs are algorithms that use individual agents and attribute changing characteristics to each one, multiple times during multiple iterations over time. This paper focuses on three philosophical aspects of ABMs as models of causal mechanisms, as generators of emergent phenomena, and as providers of explanation. Based on my discussion, I conclude that while ABMs cannot help much with causal inference, they can be viewed as etio-prognostic explanations of illness occurrence and outcome.

Keywords: Explanation, Causation, Simulation, Modelling, COVID-19.

1. Introduction

Computational modeling and simulation of real-life scenarios have become a mainstay in health research and the biosciences. My goal in this interdisciplinary paper, written from my personal perspective as physician, epidemiologist, and philosopher, is to provide an analysis of the explanatory scope of agent-based models (ABMs), one particular kind of modeling technique employed in the context of the COVID-19 pandemic (Silva et al. 2020, Cuevas 2020, Truszkowska et al. 2021, Shamil et al. 2021, Hoertel et al. 2020, Staffini et al. 2021, Kerr et al. 2021). It is not my intention to review these papers in detail; suffice it to say that they are all part of the general endeavor to tackle important population health problems posed by the COVID pandemic and have made considerable contributions to our understanding of epidemiological dynamics of this global health crisis. Instead, my discussion will focus on three philosophical aspects of ABMs as models of causal mechanisms, generators of emergent phenomena, and providers of explanation.

I will start by introducing ABMs and why they are generally considered helpful (§2). Part of their epistemological value is that they are thought to provide explanations of biological and social mechanisms (§3). One account of ABMs, featured prominently on the Columbia School of Public Health website, has ABMs as models of causal mechanisms of interactions of characteristics that may include

Argumenta 7, 1 (2021): 19-38
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.dam

impossible or unethical connections (§4) and that generate emergent phenomena (§5). The paper ends with the proposal to consider ABMs as helpful in generating etio-prognostic explanations (§6).

2. Agent-Based Models

As any other computational model, an ABM is an algorithm with inputs, computations, and outputs. In public health, ABMs are generally conceptualized as "a computational approach in which agents with a specified set of characteristics interact with each other and with their environment according to predefined rules" (Tracy, Cerdá, and Keyes 2018: 77). What exactly does that mean and why should this be helpful?

2.1 What Are ABMs?

An ABM (sometimes also called individual-based model or IBM) is a computer program that simulates changes in populations over time based on the 'behavior' of 'agents' who have a set of characteristics and 'interact' in predefined and stochastically modeled ways. This kind of simulation is often called *microsimulation* because phenomena are modeled at the micro-level (the individual agent) and results are observed at the macro-level, the level of the simulated population. Starting values and conditions for transition of agents from one state to another (for example, from non-infected to infected or from alive to dead) are defined by the programmer. Running the program will result in iterations of changes in these conditions over time. Ending conditions at the macro-level are the outcome of the model. Since the attribution of particular values to individual agents is done by randomly allocating values selected from a probability distribution with set constraints, each run of the algorithm will result in a different outcome. Multiple, oftentimes many runs need to be performed to arrive at a range of outcomes that defines an outcome distribution. The results of ABMs are non-deterministic such as those of equation-based models (EBMs). For a comparison of ABMs and EBMs, see (Van Dyke Parunak, Savit, and Riolo 1998).

Agent-based modeling is frequently used in theoretical infectious disease epidemiology (Venkatramanan et al. 2018). As outlined by Hunter and colleagues, ABMs are considered superior to EBMs (such as those that generate the now very familiar COVID-19 incidence and mortality curves) because they allow for the modelling of the behavior of individuals based on social interaction rules and a probabilistic attribution of such behaviors to the agents in a model (Hunter, Mac Namee, and Kelleher 2017). Agent-based models have to consider four major related aspects: disease, society, movement, and environment. They have to model disease-specific conditions of occurrence and duration, characteristics of the society (population) and how its members move through virtual space and interact with one another in the environment that population is situated in. The result is a highly complex representation of how population parameters change over time with regard to, e.g., disease incidence or mortality rates. Let me note right here that ABMs involve equations as well. However, the underlying equations let a set of variables undergo iterative changes over a pre-defined timeframe so that such changes over time *in each agent* contribute to an overall change at the population level.

Let us go through the published description of one ABM-based microsimulation and parse out its individual observational and inferential components.

2.2 Example: ABM of the COVID-19 Epidemic in France

Hoertel and coworkers published

> a stochastic agent-based microsimulation model of the COVID-19 epidemic in France. [They] examined the potential impact of post-lockdown measures, including physical distancing, mask-wearing and shielding individuals who are the most vulnerable to severe COVID-19 infection, on cumulative disease incidence and mortality, and on intensive care unit (ICU)-bed occupancy. While lockdown is effective in containing the viral spread, once lifted, regardless of duration, it would be unlikely to prevent a rebound. Both physical distancing and mask-wearing, although effective in slowing the epidemic and in reducing mortality, would also be ineffective in ultimately preventing ICUs from becoming overwhelmed and a subsequent second lockdown. However, these measures coupled with the shielding of vulnerable people would be associated with better outcomes, including lower mortality and maintaining an adequate ICU capacity to prevent a second lockdown (Hoertel et al. 2020: 1417).

The goal of the model was to simulate the *effect* of changing measures after the first lockdown in France such as social distancing, mask-wearing, and shielding the most vulnerable. Outcomes measures (variables) at the *population* level (macro-level) were a rebound, second lockdown, epidemic slow down, intensive care unit admission rates, mortality, as well as combinations of the above. In order to arrive at their results, investigators needed to model events at the *individual* level (micro-level), including

> 194 parameters related to French population characteristics (n = 140), social contacts (n = 33) and SARS-CoV-2 characteristics (n = 21) […]. Parameter values on population characteristics were based on data from the French National Statistical Institute (INSEE) and Santé Publique France. Parameters related to social contacts were based on prior studies (n = 11) or assumptions when no data were available (n = 22). Finally, parameters on disease characteristics were based on data from Institut Pasteur and London Imperial College, except for two unknown key parameters of the epidemic: contamination risk and proportion of undiagnosed COVID-19 cases, which were simultaneously estimated through model calibration (Hoertel et al. 2020) (quote from online material available at https://www.nature.com/articles/s41591-020-1001-6#Sec2; accessed 06/13/2021).

In essence, almost two hundred individual and population characteristics were modeled over time and the resulting changes at the population level were observed. Circling back to my tripartite goal in this paper to explore ABMs as (a) models of mechanisms, (b) generators of emergent phenomena, and (c) providers of explanation, the (a) mechanisms would be the joint changes over time among the agents of the ABM that (b) lead to certain population-based emergent phenomena, and (c) observing the model values change and results emerge would provide an explanation. The central question I ask in this paper is, an explanation of *what* exactly this might be.

2.3 Why Are ABMs Considered Helpful?

Obviously, ABMs are created for a purpose. In the general context of our current discussion, any modeler of an epidemic (pandemics included) has at least three goals. First, they want to *understand the dynamics* of the epidemic in terms of background conditions of population and environment. Thus, the first goal is to find a *causal-mechanical explanation* of why and how the infection spreads in populations. Second, modelers want to create an algorithm that allows them to *predict* how the epidemic will evolve over time. Third, modelers want to explore changes in model outcomes in response to parameter changes. In an iterative fashion, the algorithm can be modified to get closer and closer to predictions that can be confirmed or rejected by real-life data as time goes by.

I have mentioned above that one of the motivations to create ABMs is that they are considered superior to EBMs in terms of being more realistic (Hunter, Mac Namee, and Kelleher 2017). Equation-based models are simple, static, and deterministic, because they are built like a mathematical formula such as a regression equation that gives a result on a dependent variable based on the value of one or more independent variables. Once the regression equation is derived from an observational study in a certain population, any new observation can be plugged into the regression formula and a predicted value for the dependent variable can be obtained. They are static in the sense of being non-dynamic. This means that once a regression equation is created, it doesn't change. If one wants to look at other combinations of variables, different starting conditions, or changes over time, one needs to create new equations. And they are deterministic, because the value of the dependent variable is fixed once the values of the independent variables are fixed. There is not much room for "natural variation" in equation-based modelling.

Consider the following excerpt from an outline of ABMs on the website of one of the major schools of public health in the United States:

> Agent-based models are computer simulations used to study the interactions between people, things, places, and time. They are stochastic models built from the bottom up meaning individual agents (often people in epidemiology) are assigned certain attributes. The agents are programmed to behave and interact with other agents and the environment in certain ways. These interactions produce emergent effects that may differ from effects of individual agents. Agent-based modeling differs from traditional, regression-based methods in that, like systems dynamics modeling, it allows for the exploration of complex systems that display non-independence of individuals and feedback loops in causal mechanisms. It is not limited to observed data and can be used to model the counterfactual or experiments that may be impossible or unethical to conduct in the real world (https://www.publichealth.columbia.edu/research/population-health-methods/agent-based-modeling), accessed 06-04-2021).

Let me henceforth refer to this blurb as the *Columbia account of ABM* and rephrase its elements as three epistemological statements we can use as a guideline for the next sections of this paper.

Agent-based models are epistemologically helpful because they

1. enable the exploration of complex systems characterized by (among other things) non-independence of individuals and feedback loops in ***causal mechanisms***, i.e., the sequential processes of changes in agent "behavior" that

connect the initial states among agents and outcomes established at the population level;

2. support the study of interactions at the levels of people, things, places, and time between programmed behaviors of and ***interactions*** between agents that produce ***emergent effects***;
3. can explore mechanisms in ways that are ***impossible*** in observational and experimental research.

I will now turn to each one of these three epistemological benefits of ABMs in §3-5, respectively.

3. Mechanisms and Causes

3.1 Biological Mechanisms

In the basic biosciences, mechanistic views of biological processes appear to include the notions of *action* and *behavior* when it comes to the observation of changes among the mechanism's components and also the changes that occur as part of the result of the process. For example, Olaf Wolkenhauer writes that systems biologists are interested in finding out "how biological function emerges from the interactions between the components of living systems and how these emergent properties enable and constrain the behavior of those components" (Wolkenhauer 2014). First, note that Wolkenhauer says that biological function "emerges" from the interactions of components. We will come back to emergence in the next section. Second, consider this account of systems biology in light of one of the more frequently cited definitions of a mechanism in philosophy of science: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer, Darden, and Craver 2000). Taken together, the two accounts allow for the inference that at least some systems biologists see their work as identifying biological *mechanisms.*

Wolkenhauer confirms this by saying that "[t]he iterative cycle of data-driven modeling and model-driven experimentation [...] helps in identifying new mechanistic details of cell-biological processes and previously unidentified regulatory *interactions* in the system" (italics mine). Thus, another important similarity between Wolkenhauer's account of computational systems biology and Machamer, Darden, and Craver's account of mechanism is that both refer to some sort of *action*, as in "interactions" and "activities", suggesting that at least some bioscientists think that biological mechanisms are characterized by interactions and activities among the element of those mechanisms.

Let us now move from biological to population mechanics. It seems that population health scientists have a similarly mechanistic view of population health as biologists view biological processes as mechanisms. Consider, for example, this quote from the book "The Future of the Public's Health in the 21st Century" published by the Institute of Medicine (U.S.) Committee on Assuring the Health of the Public in the 21st Century (Medicine 2003): "(a)spects of discrimination might influence health through any number of mechanisms, including (socio-economic status)" (61) and "[t]here are several plausible mechanisms by which social cohesion might influence health through contextual effects" (71). These quotes raise the question how *social mechanisms* are conceptualized.

3.2 Social Mechanisms

Let us first consider who or what the elements of social mechanisms are. Stinchcombe suggests that "[m]echanisms in a theory are defined here as bits of theory about entities at a different level (e.g., individuals) than the main entities being theorized about (e.g., groups), which serve to make the higher-level theory more supple, more accurate, or more general" (Stinchcombe 1991). For our present discussion of epidemic ABMs as models of social mechanisms, the agent would be a representation of an individual person and the entirety of agents would be a representation of a social group or population. From this perspective, individuals (represented by agents in ABMs) would be the actors. In what might be the most frequently cited text on social mechanisms *as explanations*, Hedström and Svedberg (1998) confirm this when they state that their concept of social mechanism is based on four core principles, i.e., action, precision, abstraction and reduction (Hedström and Swedberg 1998). They write that

> [t]he first of these principles—explanations based on actions—means, among other things, that it is actors and not variables who do the acting. A mechanism-based explanation is not built upon mere associations between variables but always refers directly to causes and consequences of individual action oriented to the behavior of others (*ibid.*).

Are ABMs, therefore, models of social mechanisms? The following quote seems to answer in the affirmative. Conte and Paolucci write that

> [a] generative explanation of an observed social phenomenon consists of describing it in terms of the external (environmental and social) and internal (behavioral) mechanisms that generate them, rather than by inferring causes from observed covariations. This is a vital property of explanation, which cannot easily be realized otherwise. When describing agent behavior by means of other formalisms (logic-based or numeric), we describe behavior from the outside, as perceived by an observer, but do not describe the way it is generated. ABM explains (sic) behavior from within, in terms of the mechanisms that are supposed to have generated it, that is, the mechanisms that operate in the agent when s/he behaves one way or another (Conte and Paolucci 2014).

However, note that Conte and Paolucci carefully distinguish between mechanisms as *natural constituents* of the real processes the ABM is supposed to be a model of, and the *structural and functional blueprint* for agents' interactions coded into the model algorithm. They appear to see social phenomena as generated (produced) by mechanisms (external and internal) and the advantage of ABMs over other kinds of models as their capability to offer a mechanistic explanation of system behavior. Topping and colleagues make it eminently clear that the mechanisms are *built into the model.* They begin their article (about their ecological ABM model of the European brown hare) as follows:

> Agent-based models (ABMs) are gaining popularity in most scientific fields due to their ability to describe complex systems from first principles. Yet, they are also criticised for being 'black boxes' and impossible to fully understand. This is mainly due to the difficulty of testing, documenting and communicating the wealth of mechanisms built into such models (Topping, Høye, and Olesen 2010).

This view is confirmed by a group of researchers who designed an ABM on social distancing, testing, contact tracing, and quarantine on the occurrence of SARS-CoV-19 infections. Referring to multiple scenarios they modeled they write that "[t]he above scenarios are mechanistically simulated on the multi-layer network […] by allowing different interactions (between effective contacts) according to the simulated strategy" (Aleta et al. 2020). Clearly, this team stresses the point that the *simulation* is mechanistic. They do *not* say that they think that the real-life phenomena they model are mechanistic as well. However, what other reason could they have creating mechanistic models than being convinced that the modeled social and behavioral processes are mechanistic as well? Perhaps, we can paraphrase Nancy Cartwright's "no causes in, no causes out" here as "no mechanisms in, no mechanisms out" (Cartwright 1989), meaning that only if we already have mechanistic background information can we see ABMs as mechanisms. If ABMs are considered mechanistic explanations of a certain phenomenon, they explain the occurrence of the phenomenon as resulting from a mechanism by demonstrating that the phenomenon does indeed occur because of the mechanism modeled by the ABM. However, this does not yet allow for the inference that the phenomenon must be due to this mechanism. To do that, other potential mechanisms and the possibility of chance need to be ruled out, and of course the existence of the mechanism needs to be demonstrated by real world data.

3.3 Causal Mechanisms

Until now, I have tried to avoid the topic of causality because I wanted my focus to be on the role of ABMs as explaining mechanisms without reference to causation. However, some modelers talk about *causal mechanisms* when they talk about the relation between how they see causality in the world and in their models. Consider the Columbia account of ABM above: "[…] exploration of complex systems that display non-independence of individuals and feedback loops in *causal mechanisms*" (italics mine). This notion resonates with Tracy and coworkers' view that

> ABMs are well suited to the exploration of causal mechanisms given their ability to incorporate multiple interacting causes and to test competing theories about causation, thus further elucidating what we do and do not know about how a given outcome arises (Tracy, Cerdá, and Keyes 2018: 85).

It almost seems as if knowledge about mechanisms is considered crucial because it can provide knowledge about causation. As much as I agree that causes and mechanisms have a very close working relationship, they are two very different things. Indeed, Dammann has argued for a distinction between causes and mechanisms in the context of illness occurrence as separate, but closely related components of the *causation process* that represents the initial phase of illness etiology (Figure 1) (Dammann 2017). According to this account, causes initiate mechanisms that in turn culminate in clinical illness. Within this etiological scenario *all* mechanisms are causal because they link causes and their outcomes. However, this does not necessarily mean that all mechanisms must be causal. If non-causal mechanisms exist, and if ABMs can model *any* kind of mechanism, then not all mechanisms that can be represented in ABMs are causal. Therefore, any method that is supposed to extract information about *causal* mechanisms from ABMs would need to distinguish between causal and non-causal mechanisms in ABMs. On the other hand, it could be that *all* mechanisms are causal, simpliciter. We would not need to distinguish between causal and non-causal mechanisms because the latter do not exist. If all mechanisms are causal, and ABMs can model any mechanism, ABMs could be used in the exercise of generating causal-mechanical (etiological) explanations. If not, we would, again, need criteria for separating causal from non-causal mechanisms.

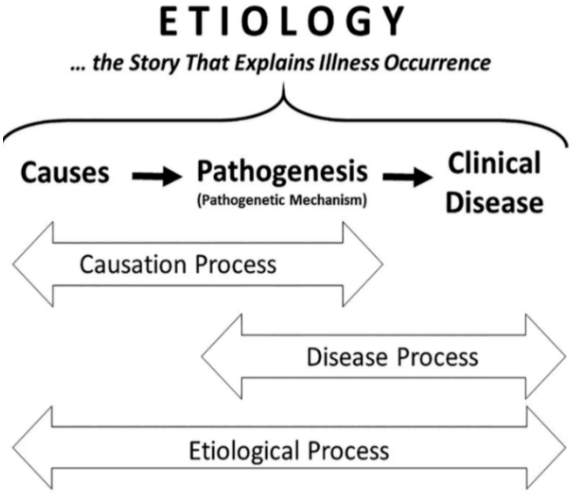


Figure 1. The etiological stance conceptualizes disease occurrence as a process. The first phase (causation process) includes causes and the subsequent pathogenetic mechanism they induce. The second phase (disease process) includes the pathogenesis and clinical disease. Knowledge about both (etiological process), combined with knowledge about the action of other contributors to the etiological process at all of its levels, can provide useful etiological explanations (reprinted with permission from Dammann 2017).

What could non-causal mechanisms look like apart from, say, non-functional mechanisms such as repetitive loops in which model parameters do not change? I am referring back to Machamer, Darden, and Craver's definition of mechanism as "entities and activities organized such that they are productive of regular changes". I take this to mean that mechanisms *produce* change. Mechanisms are the way by which causes make a difference. From this perspective, it would seem that all mechanisms are causal. Therefore, if ABMs represent mechanisms, and if all mechanisms are causal, then ABMs are representations of causal mechanisms. Does this mean that ABMs can be used as tools in causal inference?

3.4 ABMs and Causal Inference in Epidemiology

Let us assume that ABMs include causal interactions *by definition*. They are programmed to reflect a causal relationship between variables whenever one is coded to change in response to another. Indeed, this is a representation of a common causal intuition: X causes Y if Y changes whenever X changes. (I sometimes call this, somewhat informally, the light switch intuition.) It includes traditional philosophical notions of causation as regularity, difference-making, dependence, and so forth. However, I see the argument that ABMs are helpful in causal inference as being based on circular reasoning: causality in, causality out (paraphrasing Cartwright, again). ABMs cannot help with causal inference because inference is the bottom-up support of a proposition by observed data. ABMs cannot provide such data because the data they provide are top-down, generated computationally by algorithms. Yes, the *model of the algorithm* itself, e.g., the assumptions and almost 200 parameters used by Hoertel et al in our COVID-19 epidemic example above, may be based on observed data (such as disease incidence, contact frequency among agents, etc.), but the algorithm is designed to produce a result. Thus, the result is caused by the algorithm, and that causal fact does not support the notion that the underlying observed data are reflective of a causal scenario, but only the notion that the algorithm functions as a causal mechanism, and that an algorithmic causal mechanism can be interpreted as a depiction of an envisioned causal mechanism in real-life, but not as evidence supporting the inference that the modeled real-life scenario is causal or the inference that a real-life causal even exists. An algorithmic causal mechanism only shows that such mechanism has the potential to yield the modeled phenomenon. The epistemic gain is demonstrative in a theoretical way (*in silico*), but not in a practical way as in experimentation with animal models (*in vivo*). Both *in silico* and *in vivo* demonstrations confirm the possibility of a role for the mechanism in the purported etiological process, but they do not confirm that it does indeed play that role in real life scenarios.

Another caveat comes from the observation that those who argue for or against methods for causal inference via some method or another usually do so while depending on their own, implicit and often unstated intuitions about the nature of causality (Casini and Manzo 2016). What are epidemiologists' definitions of "causation"? Susser simply states that "a cause is what makes a difference" (Susser 1991). A classic paper on the counterfactual definition of *causal effect* in epidemiology includes my favorite statement "in ideal randomized experiments, association *is* causation" (Hernán 2004). My problem with this paper is, however, that it contrasts the term *causal effect* with the term *effect* because the latter is, according to the author, commonly used to mean "simply statistical association". I think the term causal effect *introduces confusion,* because there is simply no such thing as a non-causal effect. All effects are results of causal mechanisms *by definition*, although the exact mechanism itself is not always known. The more important issue here is, however, that Hernán sees the (population) definition of causal effect simply as tied to a probability differential of developing an outcome under two different exposure conditions (yes or no):

> We define the probability $\Pr[Y_a=1]$ as the proportion of subjects that would have developed the outcome Y had all subjects in the population of interest received

> exposure value a. We also refer to $\Pr[Y_a=1]$ as the risk of Y_a. The exposure has a causal effect in the population if $\Pr[Y_{a=1}=1] \neq \Pr[Y_{a=0}=1]$ (Hernán 2004: 266).

This definition strikes me as applicable to "statistical association", but by no means would I subscribe to the view that it defines "causal effect" without further explication of what Hernán *means* by "causal effect". Unless he intends to suggest that his definition *defines* causal effect. This would mean that causal effects are what epidemiologists tell us they are in a sort of metaphysically unsatisfying and somewhat patronizing way.

Let me refer briefly to an exchange from the epidemiological literature about the capability of ABMs to contribute to causal inference. Marshall and Galea have argued that ABMs "represent a promising novel approach to identify and evaluate complex causal effects" (Marshall and Galea 2015). Although they refer to causal inference in this quote and in the title of their paper, the authors seem to avoid this notion in the body of the paper and refer instead to the exploration, elucidation, and interrogation of the causal relationships modeled in an ABM. Their argument rests on the capability of ABMs to represent multiple causal interrelations (their view of a complex system):

> We argue that agent-based modeling offers an alternative and complementary approach to elucidate complex causal interdependencies that are of interest in epidemiology. Specifically, the forms of the relationships among causes (which are broadly defined here and can include agent traits as well as environments) are operationalized by the rules **Z**. The rule set consisting of functions f(), g(), and h() can include nonlinear components, including feedback loops, such that linear independence need not be assumed. By altering the rule set **Z** and running the simulation under different assumed causal relationships and processes, the effect(s) of interdependent (i.e., joint) exposures can be explored and interrogated (Marshall and Galea 2015: 96).

Marshall and Galea call the causal interrelationships they are interested in *complex*. I take it as implicit that by this they refer to complex systems, not just complicated ones. They stress the possibility to model non-linear relationships—a characteristic of complex systems. Thus, their view seems to be in keeping with the notion discussed in §5 below that the sheer complexity of interactions of agents in ABMs may give rise to emergent phenomena. More importantly, it is the intervention by the modeler (altering the rule set under different causal assumptions) that renders the ABM a helpful tool in causal exploration and interrogation, to use Marshall and Galea's terms. This view grants epistemological value to ABMs based on the possibility to *manipulate* them and explore the consequences, which resonates with interventionist accounts of causation.

One invited commentator, Ana Diez-Roux, disagrees with the notion that ABMs can help with causal inference in epidemiology (Diez Roux 2014). The following excerpt from her abstract puts her position, which I see as one point of departure for my proposal in §6 below, in a nutshell:

> As discussed by Marshall and Galea […], systems approaches are appealing because they allow explicit recognition of feedback, interference, adaptation over time, and nonlinearities. However, they differ fundamentally from the traditional approaches to causal inference used in epidemiology in that they involve creation of a virtual world. Systems modeling can help us understand the plausible

> implications of the knowledge that we have and how pieces can act together in ways that we might not have predicted. […] However, the validity of any causal conclusions derived from systems models hinges on the extent to which the models represent the fundamental dynamics relevant to the process in the real world. For this reason, systems modeling will never replace causal inference based on empirical observation. Causal inference based on empirical observation and simulation modeling serve interrelated but different purposes (Diez-Roux 2014: 100).

Of note, Diez-Roux does not say that ABMs are incapable of helping with causal inference in principle. She only says that ABM-generated models are not like epidemiological approaches to causal inference based on observed data. However, I agree with her notion that simulated data from ABMs are epistemologically inferior to observational epidemiological data simply because the underlying data are not real-world data but data generated *in silico*.

4. Interaction and Emergence

Let us now move on to the question whether the system behavior of ABMs can be reduced to the interactions among agents' characteristics and behaviors or if it is an *emergent* phenomenon. The question I am interested in is about the relationship between mechanistic explanation and emergence. In brief, if ABMs are a non-deterministic black-box and the system behavior they exhibit is truly emergent, what remains of the notion that ABMs represent causal-mechanistic explanations? What kind of causal mechanism would be explained by an ABMs whose inner workings remain in the dark and whose results are *by definition* unpredictable and surprising? (I see a similarity here to the current discussion about the transparency, explainability, and interpretability of machine learning algorithms (Roscher et al. 2020), but an exhibit of this parallel will have to wait for another day.) On the other hand, if ABMs really provide causal-mechanistic explanations we should be able to predict the phenomena they generate, which would render them non-emergent.

4.1 Emergence Defined

The classic reference on emergence, published by Jeffrey Goldstein in the first issue of the journal of the same name, defines emergence as

> the arising of novel and coherent structures, patterns, and properties during the process of self-organization in complex systems. Emergent phenomena are conceptualized as occurring on the macro level, in contrast to the micro-level components and processes out of which they arise (Goldstein 1999: 49).

Think of a complex system as having a micro level (components) and a macro level (surface). Goldstein defines emergent phenomena as (1) radically novel, (2) coherent, (3) macro-level, (4) dynamic, and (5) ostensive. *Radical novelty* refers to the fact that emergent phenomena appear at the macro level without having previously been present in the complex system under study and cannot be derived from or predicted based on knowledge about what is going on at the micro-level. *Coherence* means that emergent properties maintain "some sense of identity over time" (*ibid*.), *macro-level* means that emergence is observable at the surface level of the observed system, not the micro-level constituted by its components, *dynamic* refers to emergent phenomena as not preformed but as developing over time, and *ostensive* as being recognized by "showing themselves".

Most important for our present discussion, however, is that Goldstein sees one of the main roles of emergence in science as explanatory:

> In respect to its use in scientific explanation, the construct of emergence is appealed to when the dynamics of a system seem better understood by focusing on across-system organization rather than on the parts or properties of parts alone (Goldstein 1999: 50).

Thus, in keeping with Goldstein's characterization of emergent phenomena, although their occurrence on the macro-level is *produced* by what is going on at the micro-level, they come "out of the blue" because they do not *depend* on the behavior of individual micro-level variables (agents in ABMs) but on the overarching function of the whole system. Thus, if ABMs are truly complex (non-deterministic, non-linear) systems, they would produce emergent effects at the output level that are not predicted, or even predict*able*, by means of applying knowledge about the agents and their interactions. In contrast, these results would be ostensive occurrences that rely on the function of all interacting parts. The point here is that ABMs yield models of mechanisms that do not necessarily represent any real-world mechanism, be it biological or social mechanisms. It represents only itself, based on input conditions and probabilistic rules for agent interactions and status changes. If an ABM yields an outcome, be it emergent or expected, the occurrence of that outcome can then be explained by analyzing the workings of the modeled mechanism in silico.

What kind of mechanism consists of interactions between parts over time but is not "productive of regular changes" (per Machamer et al.'s definition) but instead to radically novel, dynamic, and ostensive phenomena? Can ABMs explain mechanisms or emergence, or both?

4.2 Weisberg: Mechanistic Explanations vs Emergence Explanations

The question whether ABMs can explain emergent phenomena is what Weisberg considers "the most controversial claim about IBMs […] Not everyone is convinced" (Weisberg 2014: 788). He quotes ecologist Joan Roughgarden as saying that she doesn't "think it's easy to discern the causation being revealed by an IBM simulation. And if we don't learn something about causation we don't learn anything scientifically important" (personal communication quoted in Weisberg 2014). (Of note, Weisberg and Roughgarden's IBMs and our ABMs are the same thing; see above.)

Weisberg suggests a distinction between explanations of emergent phenomena (mechanistic explanations) and explanations of the emergence of phenomena (emergence explanations). On his view, mechanistic explanations provide a "generalized mechanistic understanding of the dependence of higher-level properties and patterns on lower-level mechanistic factors" (Weisberg 2014: 789). I take this to mean an explanation that is based on the description of the elements of a mechanism and their interactions as being what *somehow* leads to an emergent phenomenon. He shows how certain causal graphs (relational depictions of phenomena in boxes with causal arrows between them) can depict the relationships among micro-level factors that can help generate mechanistic explanations. Interestingly, the *kind* of causal graph he chooses suggests that on his view ABMs can model *biological* mechanisms because the causal mechanism depicted in his example permits feedback loops, an important characteristic of mechanisms in biological explanations (Bechtel 2011). In contrast, the directed acyclic graphs (DAGs) that are frequently used in epidemiological causal reasoning do *not* allow feedback loops, a feature preferred in causal reasoning because the vertices can be ordered, simplifying causal argumentation immensely. No such topological order is possible in cyclic graphs (Dasgupta, Papadimitriou, and Vazirani 2008: 96).

Emergence explanations, on the other hand, would require us to provide "reductive explanations that show how emergent phenomena arise from lower-level interactions" (Weisberg 2014: 792). They would require us to clarify the *somehow* that generates an emergent phenomenon. But one main problem with both cyclic and acyclic graphs is that it is unclear *what exactly the arrows represent*. If it is true that causation is "one word, many things" and that "there are different kinds of causal relations imbedded in different kinds of systems" (Cartwright 2004: 805), the edges (arrows) between different vertices (characteristics of agents in ABMs) would potentially represent different sub-mechanisms. I read Weisberg as saying that we cannot use ABMs to provide emergence explanations unless we can specify exactly what is in each of these arrows, and I agree with him on that. On the other hand, he seems to say that ABMs can provide mechanistic explanations. Let me add that if all mechanisms are causal, I assume that Weisberg would conclude that ABMs can provide causal-mechanical explanations and I would agree with him on that as well.

I also suggest that his usage of non-DAGs to depict what ABMs model not only fits biological but also social mechanisms. Note that the Columbia account of ABMs above explicitly mentions feedback loops. Indeed, some research on COVID-19 has revealed interesting feedback loops even across scales of representation (micro-level, macro-level). For example, one computational study suggests that macro-level dynamics such as social distancing can result in micro-level changes all the way down in the genetic makeup of SARS-CoV-2 (Barrett et al. 2021).

But perhaps, at least in the context of ABMs, we shouldn't ask too much of the arrow semantics in causal graphs, for in ABMs the relationship between all agents and all their characteristics is simply a mathematical relationship, not a biological one. This brings us to the next notion reflected in the Columbia account of ABM, impossible interactions.

5. Impossible Interactions

A major motivation to use ABMs comes from their flexibility to be manipulated in ways no observational or interventional epidemiological study could be manipulated. In essence, ABMs can be used to model the "impossible" because the characteristics of agents are variables created *for* the model and *by* the model. Furthermore, there is only one kind of relationship between and among variables in ABMs, a mathematical relationship represented by stochastic functions.

Based on the findings in the systems biology and population health/sociological/ecological literature discussed in the previous sections we can postulate that ABMs are considered models of social mechanisms. Such mechanisms are modeled in ABMs by creating *interactions* between agent's characteristics among each other and between agents' and their environment's characteristics. How does this look like *inside* an ABM?

5.1 Interactions

The term *interaction* is most often used in ABMs to denote the narrowing of virtual physical space between two agents to a level at which a status change occurs in at least one of them (Winkelmann et al. 2021). Based on certain parameters, each individual agent will move through virtual space until a pre-programmed fit between a set of characteristics of two agents leads to contact and infection with a certain prespecified likelihood. At this point the status of the heretofore "uninfected" agent switches to "infected". Because such status changes are dependent on certain constellations of variables at certain timepoints, and because these constellations are derived from a whole set of characteristics assigned to agents in a stochastic fashion, these interactions and associated status changes are *not* predetermined. In this sense, ABMs are non-deterministic, and each run of the model will yield a slightly different end result. Many runs need to be performed to narrow down the probability distribution of results at the macro-level. At the population level, population wide parameters such as "infection prevalence" change from starting conditions to a different value over the duration of model run time, depending on how many individuals will be newly infected (incidence) while the model is running. Such result is sometimes considered "emergent" since it is not fully determined by model parameters in an equation-like fashion.

In the above scenario, the interaction is between two agents. Interactions can also occur between agents and the spatial environment. For example, certain areas in the virtual space can be designated as different in terms of social characteristics (e.g., high crime, low crime, no crime regions) and the likelihood of a status change of an agent (e.g., becoming the victim in a street mugging) would be different in these different regions. Moreover, agent-agent interactions could be modeled as representing just such a mugging (or not) and differ by section of the virtual space.

5.2 The Impossible

These considerations highlight one of the oft-praised advantages of ABMs, the possibility to design interactions *in any way* the modeler desires, even impossible or unethical ones. In essence, the functions of ABMs are completely devoid of the need for plausibility and ethical considerations. Nothing prevents the design of an ABM of a randomized controlled trial of the effect of COVID-19 infection on survival. Obviously, although such trial would be possible in principle, it would (luckily) never be approved by an institutional review board.

But aside from being a potential tool for modeling the unethical, another important possibility is to model mechanistic relationships across levels along the bio-psycho-social spectrum. Agent-based models can evaluate interactions among and between agents and their environments regardless of a known mechanism between, say, agents' socioeconomic background, their immune status, and their risk of SARS-CoV-2 infection. The flip side of ABMs' *inability* to provide Weisbergian emergence explanations is the benefit for the modeler to simply ignore the *somehow* expected from such explanations without sacrificing the capability of their model to provide causal-mechanical explanations.

5.3 ABMs as Multiscale Models

In epidemiological research, multilevel modeling that integrates variables across the individual, household, and community level is a common approach. Such models are called *multi-scale* or *nested* models and have become common in infectious disease modeling (Hart et al. 2020). Multiscale models have traditionally been based on integro-differential equations (IDEs), but the usage of ABMs has recently become more frequent. Such models can easily integrate the interaction between biological and behavioral processes at the level of the level of the individual and social processes at the population level.

At least some philosophers seem to feel comfortable with the idea of trans-level interaction and state that "our health is not just a metabolic response to toxins; it is about a complex social and biological interaction—a relational process or mechanism" (Parkkinen et al. 2018). Indeed, I suggest that agent-based multiscale models can provide the proposed integration of biological, behavioral, and social mechanism in a concept that Kelly, Kelly, and Russo have advocated for and called *mixed mechanisms* (Kelly, Kelly, and Russo 2014). However, I think that they can do even more: they can explore comprehensive etio-prognostic explanations of illness occurrence, development, and prognosis. Indeed, ABMs can simulate not only the joint activities of determinants of illness occurrence (causes and mechanisms) in etiological explanations (Dammann 2020), but also the joint activities of the determinants of the clinical course (disease development) and its outcomes (cure, death, or anything in between). They can even include the potential impact of etiological contributors such as *conditions* that are different from causes in non-trivial ways (Broadbent 2008) that I regret not being able to rehearse here in detail. In the next and final section, I propose that while ABMs' role in causal inference might be limited, they can provide *etio-prognostic explanations* by integrating determinants of illness occurrence (etiology) as well as determinants of disease development and outcome (prognosis).

6. ABMs as Etio-Prognostic Explanations

Above, I have rejected the idea that ABMs can help with causal inference, but support the notion that ABMs can be helpful as explanations of causal-mechanical (etiological) processes of illness occurrence. Moreover, I propose that they can help even further by simulating the trajectory of illness development and outcome. Let me begin by outlining *etiological explanations* (Dammann 2017, 2020) and what I mean by *etio-prognostic explanation.*

6.1 Etiological Explanations

In epidemiology, an obsession with causal inference abounds. The main idea seems to be that epidemiological methods can provide an apparatus that allows for causal inference based on observational epidemiological data. The underlying assumptions appear to be that observed statistical associations are not to be considered reflective of a causal relationship unless they come from ideal randomized experiments (Hernán 2004). A simple and straight forward rejection of this proposal would need to show that ideal randomized experiments do not exist. Indeed, some philosophers have offered this argument as well as other considerations that should reduce our confidence in causal inference from randomized clinical trials, the gold standard of the randomized experiment in clinical epidemiology (Worrall 2007, Cartwright 2007, 2010, Deaton and Cartwright 2016). If these arguments, which I cannot fully discuss here for reasons of space, carry any weight, there may just not be any way to reliably infer causality from epidemiological data. Instead of making causal inference the holy grail of epidemiological research, a gentler, less exclusive perspective can be taken according to which epidemiology contributes to the generation of etiological explanations, which refer to purported causes of illness, the mechanisms they initiate, and the disease (illness) that occurs. This theoretical model of illness occurrence is a process model, with causation process and disease process overlapping and jointly representing the etiological process (Figure 1). Providing such etiological explanation means providing a coherent set of hypotheses that support the observed data, explaining the occurrence of the disease and its clinical outcome (for a philosophical take on explanatory coherence in epidemiology, see Dammann 2018).

Comprehensive etiological explanations may include reference to initiators (causes), mediators, modifiers (both part of the pathogenetic mechanism), and facilitators. Causes (e.g., Sars-CoV-2 infection) are factors that initiate the mediating pathomechanism (e.g., severe inflammation in the lung) which leads to pulmonary disease, sometimes respiratory failure, and death (outcome). Modifiers in this explanation are factors that change the impact of causes and mechanisms (e.g., vaccination or social distancing), while facilitators are any biological, behavioral, or societal conditions that have an impact on the remainder of the etiological process (such as age, race, access to healthcare, and so forth). Modeling such comprehensive etiological explanation is exactly what I see multi-scale ABMs as capable of doing. They can simulate what might happen in a population given a certain constellation of characteristics that describe the interactions between initiators/causes, mediators/mechanisms, modifiers of the causation process, and facilitators/background conditions.

6.2 Etio-Prognostic Explanations

Etiological explanations are explanations that tell a cogent story of illness occurrence that is justified by reference to coherent causal and mechanistic evidence. Giving an etiological explanation means to provide a list of causes (even if the list has only one item) and mechanisms that, taken together, suffice to change the beliefs of the hitherto unconvinced about why and how the illness occurred. I think that this characterization of etiological explanations should work in both medical (single patient) settings as well as in epidemiological (population) contexts. Agent-based models that provide etiological explanations would be models of the entire etiological process from cause via mechanism to clinical disease as depicted in Figure 1. Any ABM that models COVID-19 infection incidence would provide an etiological explanation.

However, many ABMs that have been developed to model population-wide aspects of the pandemic do more: they also include estimates of hospitalizations based on estimates of illness severity, admission to intensive care, and mortality, as in the example provided above. These kinds of ABM not just explain illness occurrence but also what happens afterwards, the *prognosis* of illness. Let me offer the following table to make some potentially helpful distinctions.

<table>
<tr><th>Explanation ➔</th><th>Causal</th><th>Mechanical</th><th>Clinical</th><th>Prognostic</th></tr>
<tr><td>Explanans</td><td>Causes (risk factors)</td><td>Pathogenesis (biology)</td><td>Clinical course (signs and symptoms)</td><td>Outcome (cure, death, or anything in between)</td></tr>
<tr><td>Explanandum</td><td>Why (”roots”)</td><td>How?</td><td>Clinical presentation</td><td>Prognosis</td></tr>
<tr><td>Source of evidence</td><td>Epidemiology</td><td>Biosciences</td><td>Clinical medicine</td><td>Follow up (medicine, epidemiology)</td></tr>
<tr><td></td><td colspan="3">Etiological Explanation</td><td></td></tr>
<tr><td></td><td></td><td></td><td colspan="2">Prognostic Explanation</td></tr>
<tr><td></td><td colspan="4">Etio-Prognostic Explanation</td></tr>
</table>

Table 1. Characteristics of causal, mechanical, clinical, and prognostic explanations.

Of note, the “intended *explicandum* [of scientific explanations] is, very roughly, explanations of *why* things happen, where the ‘things’ in question can be either particular events or something more general—e.g., regularities or repeatable patterns in nature” (Woodward and Ross 2021). I am aware that explaining *why* something happens is a very different thing than explaining its *consequences*. Indeed, such an explanation would probably not be considered *scientific*. However, a slight change of perspective might allow us to reintroduce science through the backdoor. We could say that what happens after the initial occurrence of illness is just the *occurrence* of aspects of disease development and outcome. Thus, the prognostic part of etio-prognostic explanations can be viewed as providing a plain old etiological explanation. This way, one could see prognostic explanations as scientific, i.e., by recognizing them as etiological explanations of a different target entity.

However, I am interested in the mere *practical* usefulness of explanations of illness occurrence and outcome. I prefer looking at ABMs as providing a pragmatic kind of explanation, which is simply helpful by illuminating both the *etiology and prognosis* of illness. This is exactly what we expect from ABMs in the context of the COVID-pandemic: explanations why and how illness occurrence patterns arise at the population level, how they evolve, and what their consequences are.

7. Conclusion

In this paper, I have discussed the epistemological characteristics of ABMs, one type of simulation model used in the context of the COVID-19 pandemic. In contrast to equation-based models, ABMs are algorithms that use individual agents and attribute changing characteristics to each one, multiple times during multiple iterations over time. Based on my discussion, I conclude that ABMs can explain causal mechanisms but cannot provide emergence explanations, because they cannot provide information about exactly why low-level phenomena give rise to those emergent phenomena. This is also one reason why I believe that ABMs cannot help with causal inference. Another reason is that ABMs do not reflect real-world processes but the causal-mechanical intuitions of the modeler. On the other hand, ABMs can integrate "impossible" multi-scale interactions between initiators, mediators, moderators, and conditions, and may be useful as comprehensive etio-prognostic explanations of illness occurrence and outcome.

References

Aleta, A., Martin-Corral, D., Piontti, Y. Pastore, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N.E., Halloran, M.E., Longini, I.M. Jr., Merler, S., Pentland, A., Vespignani, A., Moro, E., and Moreno, Y. 2020, "Modelling the Impact of Testing, Contact Tracing and Household Quarantine on Second Waves of COVID-19", *Nature Human Behaviour*, 4, 9, 964-71, doi: 10.1038/s41562-020-0931-9.

Barrett, C., Bura, A.C., He, Q., Huang, F.W., Li, T.J.X., Waterman, M.S., and Reidys, C.M. 2021, "Multiscale Feedback Loops in SARS-CoV-2 Viral Evolution", *Journal of Computational Biology*, 28, 3, 248-56. doi: 10.1089/cmb.2020.0343.

Bechtel, W. 2011, "Mechanism and Biological Explanation", *Philosophy of Science*, 78, 533-57.

Broadbent, A. 2008, "The Difference between Cause and Condition", *Proceedings of the Aristotelian Society*, 108, 1, pt 3, 355-64.

Cartwright, N. 1989, *Nature's Capacities and Their Measurement*, Oxford-New York: Clarendon Press.

Cartwright, N. 2004, "Causation: One Word, Many Things", *Philosophy of Science*, 71, 805-19.

Cartwright, N. 2007, "Are RCTs the Gold Standard?", *Biosocieties*, 1, 11-20.

Cartwright, N. 2010, "What Are Randomised Controlled Trials Good For?", *Philosophical Studies*, 147, 1, 59-70.

Casini, L., and Manzo, G. 2016, "Agent-Based Models and Causality: A Methodological Appraisal", *The IAS Working Paper Series*, 7, 1-80.

Conte, R., and Paolucci, M. 2014, "On Agent-Based Modeling and Computational Social Science", *Frontiers in Psychology*, 5, doi: 10.3389/fpsyg.2014.00668.

Cuevas, E. 2020, "An Agent-Based Model to Evaluate the COVID-19 Transmission Risks in Facilities", *Computers in Biology and Medicine*, 121, doi: 10.1016/j.compbiomed.2020.103827.

Dammann, O. 2017, "The Etiological Stance: Explaining Illness Occurrence", *Perspectives in Biology and Medicine*, 60, 2, 151-65, doi: 10.1353/pbm.2017.0025.

Dammann, O. 2018, "Hill's Heuristics and Explanatory Coherentism in Epidemiology", *American Journal of*Epidemiology, 187, 1, 1-6, doi: 10.1093/aje/kwx216.

Dammann, O. 2020, *Etiological Explanations: Illness Causation Theory*, Boca Raton: CRC Press.

Dasgupta, S., Papadimitriou, C.H., and Vazirani, U.V. 2008, *Algorithms*, Boston: McGraw-Hill Higher Education.

Deaton, A., and Cartwright, N. 2016, "Understanding and Misunderstanding Randomized Controlled Trials", *National Bureau of Economic Research Working Paper Series*, No. 22595, doi: 10.3386/w22595.

Diez Roux, A.V. 2014, "Invited Commentary: The Virtual Epidemiologist—Promise and Peril", *American Journal of Epidemiology*, 181, 2, 100-102, doi: 10.1093/aje/kwu270.

Goldstein, J. 1999, "Emergence as a Construct: History and Issues", *Emergence*, 1, 1, 49-72, doi: 10.1207/s15327000em0101_4.

Hart, W.S., Maini, P.K., Yates, C.A., and Thompson, R.N. 2020, "A Theoretical Framework for Transitioning from Patient-Level to Population-Scale Epidemiological Dynamics: Influenza A as a Case Study", *Journal of The Royal Society Interface*, 17, 166, doi: 10.1098/rsif.2020.0230.

Hedström, P. and Swedberg, R. 1998, *Social Mechanisms: An Analytical Approach to Social Theory*, in Elster, J. and Hernes, G. (eds.), *Studies in Rationality and Social Change*, Cambridge-New York: Cambridge University Press.

Hernán, M.A. 2004, "A Definition of Causal Effect for Epidemiological Research", *Journal of Epidemiology and Community Health*, 58, 4, 265-71.

Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Sánchez Rico, M., Limosin, F., and Leleu, H. 2020, "A Stochastic Agent-Based Model of the SARS-CoV-2 Epidemic in France", *Nature Medicine*, 26, 9, 1417-21, doi: 10.1038/s41591-020-1001-6.

Hunter, E., Mac Namee, B., and Kelleher, J.D. 2017, "A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology", *Journal of Artificial Societies and Social Simulation*, 20, 3, doi: 10.18564/jasss.3414.

Kelly, M. P., Kelly, R.S., and Russo, F. 2014, "The Integration of Social, Behavioral, and Biological Mechanisms in Models of Pathogenesis", *Perspectives in Biology and Medicine*, 57, 3, 308-28, doi: 10.1353/pbm.2014.0026.

Kerr, C.C., Robyn, M., Stuart, D.M., Romesh, G.A., Rosenfeld, K., Hart, G.R., Núñez, R.C., Cohen, J.A., Selvaraj, P., Hagedorn, B., George, L., Jastrzębski, M., Izzo, A., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S., Bennette, C.S., Wagner, B., Chang, S., Oron, A.P., Wenger, E., Panovska-Griffiths, J., Famulare, M., and Klein, D.J. 2021, "Covasim: An Agent-Based Model of COVID-19 Dynamics and Interventions", *medRXiv*, https://www.medrxiv.org/content/10.1101/2020.05.10.20097469v3

Machamer, P.K., Darden, L. and Craver, C.F. 2000, "Thinking About Mechanisms", *Philosophy of Science*, 67, 1-25.

Marshall, B.D., and Galea, S. 2015, "Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology", *American Journal of Epidemiology*, 181, 2, 92-99, doi: 10.1093/aje/kwu274.

Medicine, Institute of 2003, *The Future of the Public's Health in the 21st Century*.

Parkkinen, V., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M.P., Norell, C., Russo, F., Shaw, B., and Williamson, J. 2018, *Evaluating Evidence of Mechanisms in Medicine*, Cham: Springer.

Roscher, R., Bohn, B., Duarte, M.F., and Garcke, J. 2020, "Explainable Machine Learning for Scientific Insights and Discoveries", *IEEE Access*, 8, 42200-42216, doi: 10.1109/access.2020.2976199.

Shamil, M.S., Farheen, F., Ibtehaz, N., Khan, I.M., and Sohel Rahman, M. 2021, "An Agent-Based Modeling of COVID-19: Validation, Analysis, and Recommendations", *Cognitive Computation*, doi: 10.1007/s12559-020-09801-w.

Silva, P.C.L., Batista, P.V.C., Lima, H.S., Alves, M.A., Guimarães, F.G., and Silva, R.C.P. 2020, "COVID-ABS: An Agent-Based Model of COVID-19 Epidemic to

Simulate Health and Economic Effects of Social Distancing Interventions", *Chaos, Solitons & Fractals* 139, doi: 10.1016/j.chaos.2020.110088.

Staffini, A., Svensson, A.K., Chung, U., and Svensson, T. 2021, "An Agent-Based Model of the Local Spread of SARS-CoV-2: Modeling Study", *JMIR Medical Informatics*, 9, 4, doi: 10.2196/24192.

Stinchcombe, A.L. 1991, "The Conditions of Fruitfulness of Theorizing About Mechanisms in Social Science", *Philosophy of the Social Sciences*, 21, 3, 367-88, doi: 10.1177/004839319102100305.

Susser, M. 1991, "What is a Cause and How Do We Know One? A Grammar for Pragmatic Epidemiology", *Am J Epidemiol*, 133, 635-48.

Topping, C.J., Høye, T.T., and Olesen, C.R. 2010, "Opening the Black Box—Development, Testing and Documentation of a Mechanistically Rich Agent-Based Model", *Ecological Modelling*, 221, 2, 245-55, doi: 10.1016/j.ecolmodel.2009.09.014.

Tracy, M., Cerdá, M., and Keyes, K.M. 2018, "Agent-Based Modeling in Public Health: Current Applications and Future Directions", *Annual Review of Public Health*, 39, 1, 77-94, doi: 10.1146/annurev-publhealth-040617-014317.

Truszkowska, A., Behring, B., Hasanyan, J., Zino, L., Butail, S., Caroppo, E., Jiang, Z., Rizzo, A., and Porfiri, M. 2021, "High-Resolution Agent-Based Modeling of COVID-19 Spreading in a Small Town", *Advanced Theory and Simulations*, 4, 3, doi: 10.1002/adts.202000277.

Van Dyke Parunak, H., Savit, R., and Riolo, R.L. 1998, *Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and Users' Guide*, in Sichman, J.S., Conte, R., and Gilbert, N. (eds.), *Multi-Agent Systems and Agent-Based Simulation*, Berlin: Springer, 10-25.

Venkatramanan, S., Lewis, B., Chen, J., Higdon, D., Vullikanti, A., and Marathe, M. 2018, "Using Data-Driven Agent-Based Models for Forecasting Emerging Infectious Diseases", *Epidemics*, 22, 43-49, doi: 10.1016/j.epidem.2017.02.010.

Weisberg, M. 2014, "Understanding the Emergence of Population Behavior in Individual-Based Models", *Philosophy of Science*, 81, 5, 785-97, doi: 10.1086/677405.

Winkelmann, S., Zonker, J., Schütte, C., and Conrad, N.D. 2021, "Mathematical Modeling of Spatio-Temporal Population Dynamics and Application to Epidemic Spreading", *Mathematical Biosciences*, 336, doi: 10.1016/j.mbs.2021.108619.

Wolkenhauer, O. 2014, "Why Model?", *Frontiers in Physiology*, 5, 21, doi: 10.3389/fphys.2014.00021.

Woodward, J. and Ross, L. 2021, "Scientific Explanation", in *The Stanford Encyclopedia of Philosophy*, Zalta, E.N. (ed.).

Worrall, J. 2007, "Why There's no Cause to Randomize", *British Journal for the Philosophy of Science*, 58, 3, 451-88, doi: 10.1093/Bjps/Axm024.

KISSing in the Time of COVID-19: Some Lessons for Model Choice

Till Grüne-Yanoff

KTH Royal Institute of Technology, Stockholm

Abstract

I present and analyze the case of COVID-19 modeling at the *Public Health Agency of Sweden* (FoHM) between February 2020 and May 2021. The analysis casts the case as a decision problem: modelers choose from a strategically prepared menu that model which they have reasons to believe will best serve their current purpose. Specifically, I argue that the model choice at FoHM concerned a trade-off between model-target similarity and model simplicity. Five reasons for choosing to engage in such a trade-off are discussed: lack of information, avoiding overfitting, avoiding fuzzy modularity, maintaining good communication, and facilitating error avoidance and detection. I conclude that the case illustrates that model simplicity is an epistemically important principle.

Keywords: Modelling, Methodology, Similarity, Simplicity, Epistemic virtues.

1. Introduction

The epidemiological modelling toolbox has grown considerably over the last twenty years. The Public Health Agency of Sweden (FoHM) is a good illustration of that: it has systematically developed its menu of mathematical and computational modeling tools. But constructing a menu also forces a choice, and this is the focus of my case study in this paper. When Covid-19 came to Sweden, how did FoHM modelers choose their modeling tools from those menu options?

That is an interesting story in its own right, which I will sketch here, but my underlying interest is to identify the *reasons* for this choice. For this purpose, I first rehash in section 2 the discussion of two opposing desiderata, first of model-target similarity, and second of model simplicity, as exemplified by the KISS principle: "Keep It Simple Stupid". I argue that these are the relevant criteria for my case, by showing that the different options contained in the toolbox indeed differ mainly in how much they simplify and how many parameters they include (section 3.1). I then recount how the COVID-19 models at FoHM were chosen (section 3.2). In section 4, I discuss five epistemic reasons for choosing the simpler kind of model, in effect sacrificing a certain degree of potentially higher model-target similarity for the sake of more model simplicity. Section 5 concludes.

Argumenta 7, 1 (2021): 39-55
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.gru

2. Modeling Between Similarity and Simplicity

At the heart of every methodological question about models lies a decision. The modeler pursues a certain goal—for example, predicting a future event based on current data, or explaining a current phenomenon with available theory—and then needs to choose which available model to employ for this purpose. This decision depends on other, previous choices—whether to model at all, and what models to make available for oneself—but in this paper I want to focus on the choice between available models.

Models can be assessed according to many different criteria, and some of these might trade off on each other (Levins 1966, Matthewson and Weisberg 2009, Elliott and McKaughan 2014). Even if there is no general trade-off, such desiderata can come into conflict under certain conditions. This is the case with similarity and simplicity, the two model virtues I focus on in this paper.

The similarity desideratum derives from the idea that models represent specific targets. Targets might either be actual or non-actual things; although scientists are interested in them, they represent them with models and then investigate the models in their stead. This might be because the target is not accessible or cannot be manipulated, be this for physical, economic or legal reasons. For example, cosmologists model black holes because they currently cannot be accessed or manipulated; pharmacologists experiment with rat models, because experimenting on humans is very restricted; economists use macroeconomic models because experimenting with the interest rate could have grave economic consequences.

This rationale for modelling implies two important consequences. First, if models are supposed to function as stand-ins or surrogates of targets for one of the above reasons, then it is important that they are similar to these targets. The less similar a model is to its target, so it seems from this perspective, the more difficult it is to justify its use in its target's stead. Similarity seems to emerge as a prime model desideratum from these considerations, and many authors have indeed defended such a view, with qualifications regarding background theory and modeler's purpose (Giere 1988, Weisberg 2013).[1]

Second, however, if models are employed because of one of the obstacles that the target poses for a direct investigation, then any successful model must differ from the target at least with respect to that obstacle. To be useful, a black hole model must be accessible and manipulable; it must be legal to experiment on the modal organism; and the manipulation of the Macro-model must not put millions into the poorhouse. Consequently, models must be to some degree different from their targets; demanding full similarity or identity between model and target would defeat the very rationale of using models in the first place.

Once one admits that much however, the question arises *how much* similarity to demand between model and target. Here views differ considerably. Some argue for similarity to a large extent: "Fruitful models […] share many, and do not fail to share too many, features [with their targets] that are thought to be salient by the scientific community" (Weisberg 2013: 155). Others, however, have argued for sacrificing some degree of similarity for the sake of simplicity (Paola and

[1] Similarity was once thought to be a vacuous concept (Goodman 1972). The more recent literature offers a number of operationalizations (e.g. Weisberg 2013), even if these remain controversial (Parker 2015).

Leeder 2011). These two desiderata trade off on each other if the target has a high degree of complexity: stressing similarity would then make the model complex, while stressing simplicity would keep its complexity comparatively lower, at the cost of reduced similarity.[2]

Model builders might take inspiration from engineers, who widely accept simplicity as a design principle, for at least two reasons. First, the more complex the device, the more difficult it is to control; keeping design simple thus allows, *ceteris paribus*, better control. Second, design defects are better remedied by basic redesign than by superficial modification, as this avoids conservative *ad hoc* maneuvers. Keeping devices simple forces such early redesign in the face of defects. In design, these considerations widely became known as the KISS principle (Rich 1995).

Mutatis mutandis, the KISS design principle has also been applied to modelling choice. In the computer simulation community, however, criticism of KISS grew with increasing computational power available to modelers. Why, these authors asked, should one sacrifice any degree of similarity if increased computational capacities allowed the construction and analysis of models of hitherto unreachable detail-richness? Some authors even formulated a counter principle to KISS, which they termed KIDS: "Keep It Descriptive Stupid", thus explicitly endorsing a maximization of similarity:

> The KIDS approach starts with a model which relates as strongly to the target domain as possible, but does not ensure that the models are "elegant". Before the advent of cheap computational power, it was only possible to get any results out of analytic (and hence relatively simple) models; this made the KIDS approach infeasible (Edmonds and Moss 2004: 142).

This perspective on the similarity-simplicity trade-off will be central for this paper. Why, at a time of newly-won and still increasing technical feasibility, would modelers want to give up any potential similarity between model and target for the sake of keeping the model simple? With this question I turn to the case of COVID-19 modeling at the Public Health Agency of Sweden in 2020.

3. The Case: COVID-19 Modeling in Sweden

The Public Health Agency of Sweden (*Folkhälsomyndigheten*, FoHM) is a Swedish government agency with national responsibility for public health. It was formed in 2014 by a merger of the Swedish National Institute of Public Health (*Folkhälsoinstitutet*) and the Swedish Institute for Communicable Disease Control (*Smittskyddsinstitutet*, SMI). It has about 600 employees in six departments. Its task is to produce and disseminate scientifically sound knowledge that promotes health and prevents diseases and injuries. Its target groups are the national government, other state authorities, regions, and various interest groups (FoHM 2021c).

Epidemiological modeling at FoHM is performed at the Analysis Unit, which is part of the Department of Public Health Analysis and Data

[2] There are multiple notions of simplicity relevant for model choice. Rochefort-Maranda (2016), for example, distinguishes between parametric, theoretical, computational, epistemic, and dimensional simplicity. I will focus on parametric simplicity in this article.

Management. Since its inception, the unit has been headed by Dr. Lisa Brouwers, with a staff of 20, out of which 4-5 work with epidemiological modelling. Lisa received her PhD in Computer and Systems Sciences from Stockholm University in 2005. The title of her thesis was *Microsimulation Models for Disaster Policy Making*. Since 2004, she had been associated with the SMI, one of the predecessors of FoHM.

According to Brouwers, mathematical modelling—beyond statistical regression analysis—had not been practiced at FoHM until the early 2000s. This reflects the then-state of epidemiology more broadly: many epidemiologists in the early years of the millennium did not consider analytical models as part of their toolkit (Chubb and Jacobsen 2010; for a survey amongst epidemiologists about mathematical modelling, see Hejblum et al. 2011). Brouwers was hired at SMI into a project that aimed changing that.[3]

> I have been lucky to have managers who were interested and saw the relevance of modeling, so I have had the opportunity to over the years strengthen and form the modeling capacity within SMI, and then later on at FoHM (Brouwers interview 2021: 2).

Brouwers was tasked with implementing a long-term modelling strategy at FoHM: "it has been my responsibility to make sure we have had competence in modeling" (Brouwers interview 2021: 2). In particular, this involved constructing and maintaining a number of distinct model frameworks for epidemic modelling—a modeling toolbox—and developing staff competences in maintaining and applying them:

> What we had in mind was to have different types of models available or quite ready to deploy when we need them. The *MicroSim* model was one option, the *SEIR* models or variations of them, and then more statistical models. But maybe most of all we made sure that we had staff—competence—who can program such models, as well as have the ability to decide which models to start working with and when (Brouwers interview 2021: 2).

FoHM thus pursued a strategy of systematically developing and cultivating a menu of modelling tools, from which model choice for different epidemiological purposes could be made. Before describing the choice itself, it is worthwhile detailing what this menu actually consists of.

3.1 The Modelling Toolbox at FoHM

Compartmental models are some of the most commonly used models in infectious disease epidemiology, and various versions of these are also part of FoHM's toolbox. Its most popular version, the *SIR* model, consists of three compartments: *S* (susceptible) for individuals at risk of infection, *I* (infectious) for individuals currently infected, and *R* (recovered) for individuals who recovered from the infection and have immunity. Every individual in a population is assigned to one

[3] The project was headed by Johan Giesecke, state epidemiologist of Sweden from 1995 to 2005. During the pandemic, FoHM contracted again Giesecke, by then professor emeritus, to "support the Unit of analysis in their modelling and analysis of COVID-19, at a maximum of 800 h in 2020" (Karlsten 2020, my translation).

of the three compartments. Within each compartment, individuals are assumed to have the same properties and act in the same way. Individuals may progress between compartments according to predefined flow patterns. In the *SIR* model, for example, individuals progress from *S* to *I* to R. To model a specific epidemic in a particular population, one quantifies the proportion of the population located in each compartment at a specific time and assigns values to the rates of flow between compartments. The *SIR* model is commonly run with ordinary differential equations, which are deterministic. Alternatively, the parameters specifying the flow rates can be expressed as probability distributions to better capture the uncertainty of the estimates.

The *SIR* model can be varied by changing the number of compartments—either by expanding it (e.g. adding an 'exposed' compartment in a *SEIR* models, or connecting *R* back to *S* in a *SIRS* model, where immunity lasts only for a short period of time) or by contracting it (e.g. to a simple logistic *SI* model, or a *SIS* model where there is no immunity). Furthermore, compartmental models have occasionally included seasonally dependent flow rates, diffusion constants to model spatial distribution of the infected, vital statistics like births and deaths, age distributions and vaccination status. What all these variants share is the assumption that the transition rate between *S* and *I* is determined by the average number of contacts per person per time unit, in conjunction with the probability of disease transmission in a contact between a susceptible and an infectious subject. Individuals are assumed to mix homogeneously: their contact rates are assumed to be independent of their individual identities. Typical uses of such models include the prediction of disease spread, total number infected, epidemic duration or the infection's peak.

FoHM developed such models. Specifically, at the beginning of COVID-19, it used a deterministic *SEIR* model with compartments *E* for exposed and a distinction between I_r for reported and I_o for unreported infections. The flow pattern of this model was: *S* -> *E* -> $\{I_r, I_o\}$ -> *R* (FoHM 2020b). This model was later expanded to include two R-compartments: R_1 in which an individual still can test positive on a PCR test, and a second compartment R_2, in which an individual no longer tests positive on a PCR test (FoHM 2020d). In a third analysis, called *VirSim*, the initial *SEIR* model was modified to contain three separate age cohorts (0-19, 20-69, 70+) (FoHM 2020c). This model had already been developed at SMI ten years earlier, under participation of Brouwers (Fasth et al 2010). Until June 2021, all COVID-19 analyses published by FoHM relied on some variant of these compartment models.

However, these models were not the only ones in FoHM's toolbox. To the contrary, Lisa Brouwers' early research work concentrated on another kind of model, in which individuals and their contacts are represented explicitly and heterogeneously. Such models are often called agent-based models or microsimulations. Brouwers developed her first microsimulation model, *MicroPox*, as part of her PhD (Brouwers 2005). The model is a microsimulation model, representing all 8,861,393 Swedes (the size of the Swedish population when the data set was collected). A unique feature of the model is that it uses government census data on where each person works, who the person works with, and who the person lives with. This makes it possible to extract a network of contacts that shows the professional and family contacts. These contacts are depicted deterministically. A day in the simulation model is divided into day and night. In the first hour of the day, people with a job go to work. If people are unemployed or retired, they stay at home. Since school and kindergarten data were not available, the model uses a proxy based on age and physical distance. Everyone returns home after work and sleeps there with their families. The model gives transmission probabilities for each of these locations (Brouwers 2005, Brouwers and Liljeros 2005).

Brouwers illustrated the model use at the hand of a number of simulated smallpox epidemics. This was motivated by the knowledge that smallpox spreads mainly through close contacts, and that therefore the contact network is of greater importance than it would be for a highly contagious disease like measles. The purpose of these modelling efforts was to create a tool for testing the effects of intervention policies, including mass vaccination, targeted vaccination, isolation and social distancing. Initially focused on smallpox, the model, renamed *MicroSim*, was modified to support simulations of pandemic influenza in 2006 (Brouwers et al. 2009a). Specifically, *MicroSim* was used to estimate the economic consequences of reduced absenteeism through a sufficiently strong H1N1 influence vaccination campaign (Brouwers et al. 2009b).

Models like *MicroPox* or *MicroSim* are costly to maintain, both in terms of the regularly needed census data updates, as well as in terms of the staff competences and worktime required for their maintenance.[4] Despite these costs, FoHM kept these models in their toolbox:

> They are kind of, or at least they were before COVID started, maintained up to date. I have had persons working with what we call *MicroSim* [...] *MicroSim* is quite maintained, documented, and possible to run. In other words, yes, it would be possibly without too much trouble to get it started again (Brouwers interview 2021: 1).

Not all agent-based models are as complex, detail-rich and data-intensive as *MicroPox* or *MicroSim*, however. Other agent-based models might still represent individuals and their contacts explicitly and heterogeneously, but focus on a smaller population or avoid reliance on census data altogether.

As an example, consider Burke et al. (2006), who simulated a single initial infected person attack on a town network of either 6,000 or 50,000 people. Town networks either consist of one town, a ring of six towns, or a 'hub' with four 'spokes.' Each town consists of households of up to seven persons, one workplace, and one school. All towns share a hospital. Each space is represented as a grid, so that each cell in the grid has eight neighbors. Agents are distinguished by type (child, health care worker, commuter) by family ID and by infectious status. Each 'day,' agents visit spaces according to their type, and then return home. On the first day of the simulation, the position in schools and workplaces is randomly assigned, but after that, agents remember their positions. During the day, agents interact with all of their immediate neighbors: 10 times at home, 7 times at work, and 15 times in the hospital. After each interaction, they move positions to the first free cell in their neighborhood. Homogeneous mixing is thus completely eschewed; instead, agents interact in a number of dynamic neighborhoods. These models represent some recognizable "town-properties" without representing any

[4] "I think that to maintain a microsimulation or agent-based model you either need to have a large group of modelers so you can have one or two modelers working with it part-time every year, or you need a specific interest to drive the project yourself as head of the modeling group, or maybe some other collaborations with a university" (Brouwers interview 2021: 3).

actual town or drawing on any data from such actual towns (For a more detailed discussion, see Grüne-Yanoff 2021).
These intermediate agent-based models are also part of FoHM's toolbox.

> We use these kinds of models as well. […] it is more of a tool to study network effects of different phenomena. […] When we implement our *SEIR* models in different age groups they become so complex that it is easier to implement the models as agents, i.e. each agent or each version of the *SEIR* structure is an implementation of an agent. So, we are kind of using it but not conceptually as an agent-based model. Nevertheless, we are considering the whole spectra of different kinds of models (Brouwers interview 2021: 3).

To summarize, FoHM for almost 20 years has developed a modelling toolbox, consisting of "the whole spectra of models" to suit various modelling purposes. This toolbox included both compartmental models, census-based agent-based models and more abstract agent-based models. Given this menu of available tools, it is interesting to see how FoHM actually chose its models for investigating COVID-19, when the epidemic came to Sweden.

3.2 Choosing the Covid-19 Model February 2020-May 2021

In February 2020, Brouwers recalls, her boss first suggested that they start looking at their models and get prepared to assist the Swedish regions in making predictions in terms of hospital and care needs. "Out of curiosity", Brouwers and her team also tried to see if they could fit *SEIR* models to the very early data available from Wuhan. However,

> quite early we realized that we could not do it because the *SEIR* models overshoot by predicting huge outbreaks. We of course understood that this happened because there must be an unreported fraction of infected persons [whose] size we did not know. Also, we thought that it is probably not a totally homogenous spread. In other words, there was still so much we did not know about how the virus was spread—how infectious it was. Simply taking the data from Wuhan and implement it in models for Sweden would render enormous outbreaks that were not realistic, because otherwise we would have seen the local outbreaks spread faster, probably, from Wuhan to the rest of China at that time (Brouwers interview 2021: 4).

This failure at replicating the Wuhan data with compartmental models led to two decisions. First, in order to estimate the burden for hospitals in the regions, FoHM would initially not try to replicate the data or model the dynamic, even in an *SEIR* model, for a prediction of the outbreak in Sweden. Instead, they used prototypical epi-curves where they beforehand decided how many would be infected, in order to sketch a realistic worst-case scenario for the regions, without modelling any transmission. This led to the first report on March 20th (FoHM 2020a).

> We thought that this could be used to help the regions in Sweden by providing an answer [to] the following question: If it in each region of Sweden would become as bad as it was in Wuhan, then what would the need of hospital beds be? Divided into ICU-beds and ordinary hospital beds. So we constructed outbreaks with a clinical attack rate of the same magnitude as in Wuhan, 1%, for each region in Sweden, using a simple *SIR*-model (Brouwers interview 2021: 4).

The second decision was that the explicit modelling of transmission with any model should wait until relevant data was available about the pandemic in Sweden—in particular on the fraction of unreported cases. In the so-called *Gloria*-studies, FoHM in collaboration with the Swedish army tested population samples in different regions for COVID-19 and from this concluded that 98.7% of all infections go unreported (FoHM 2020e).

> At that time, we had that piece of puzzle we missed previously, so with this modelling report we switched from just doing some prototypical to try to make a realistic representation of the dynamics in the region where we had that information we were previously lacking (Brouwers interview 2021: 7).

With that information they built the first *SEIR* model with separate compartments I_r for reported and I_o for unreported infections. Because the infection rate of unreported cases might differ from the reported ones, they modelled three scenarios with different infection rates—one identical to, another at 55%, and one at 11% of those in I_r. (FoHM 2020b). The purpose of this model was to estimate when the infection would peak in Stockholm, and how many infected were to be expected until the end of April.

When the results of this study were presented at a press conference on April 21st, deputy state epidemiologist Anders Wallensten explained the model's assumption with an illustration, saying that "there is about one confirmed case of COVID-19 out of 1000 cases in total". The figure caused confusion. As a journalist pointed out, at that time about 6,000 infections were confirmed in Stockholm alone—would this imply that almost six million were actually infected (Stockholm region has less than 2,5 million inhabitants). Brouwers and her team stated that the figure was incorrect—a fact that the press reported probably more than any other results from FoHM's analysis unit.

Nevertheless, Brouwers two days later could also see a positive side of this hiccup: "It's almost lucky that the mistake was so obvious, it could have been much more subtle" and hard to find, she said in an interview with *Aftonbladet.*[5] Before the publicly accessible code was taken down, just after the 21/4 press conference, five members of the public had already written to the programmer about the same error. The model was reprogrammed, and the report adjusted accordingly.

FoHM has performed a large number of model studies of Covid-19 since, including projections of rising infection rates due to increased summer travel (FoHM 2020c), estimates of infections rates in some of the Swedish regions until early 2021 (FoHM 2020d) and scenarios for future developments of the pandemic into summer 2021 (FoHM 2020f, FoHM 2021a and 2021b). All of these reports are based on compartmental models, the later ones specifically on FoHM's *VirSim.* Not a single study was performed with agent-based tools, specifically not *MicroSim*.

The main reason for that was a lack of relevant data, especially about the basic epidemiology and the medical features of the disease. Comparing it to the influenza simulation they did with *MicroSim* before, Brouwers argued that

[5] "Det var nästan tur att felet vår så uppenbart, det kunde ha varit klart mer subtilt" (Karlsson 2020).

> When we model the flu, we have quite a lot of data, but we still have to make a lot of assumptions regarding how many days after infection individuals' peak their infectiousness, how do people act when they are asymptomatically infected, etc. But for COVID we had no such information at all, and therefore we decided that moving to a *MicroSim* model with so much uncertainty is not worth it. It would not have been wise. Further, no one pushed for such a *MicroSim* model either (Brouwers interview 2021: 6).

This does not exclude that FoHM will soon begin making use of *MicroSim*, however:

> After the summer [2021] it could be the case that we start doing more network and maybe agent-based modeling to look at the spread within certain groups, or between certain groups, and the rest of the population where you have a quite heterogeneous vaccination coverage in the population—pockets with the risk of infection put among young people and also some groups that have a lower vaccination coverage. We are discussing this, but we have not started this modeling yet (Brouwers interview 2021: 5).

There are legitimate purposes for which agent-based simulations might be used in modelling COVID-19; yet these purposes have not been the immediate pressing ones during the early and current stages of the epidemic—at least not under the given circumstances. How such considerations of purpose and circumstance constitute reasons for choosing certain options from the modelling toolbox will be my focus in the next section.

4. KISS Despite Technical Feasibility: Reasons for Model Choice

Modelers at FoHM systematically build a toolbox comprising the whole spectrum of epidemiological model types. They did so because they believed that each of these model types had their own advantages that made them a best choice for certain purposes, under certain conditions. The long-term strategy was to provide the technical means to quickly apply the best model to a host of possible eventualities. When faced with the COVID-19 outbreak in Sweden, modelers at FoHM chose compartmental models over agent-based models, even though the latter were technically feasible and available. In this section I investigate the reasons for this choice—both those explicitly considered by FoHM staff, as well as implicit ones that could justify such trade-off decisions.

This choice is philosophically interesting, because it exemplifies the trade-off between simplicity and similarity sketched in section 2. The compartmental models weren't simply better than the agent-based models; to the contrary, the FoHM modelers explicitly acknowledge that the compartmental models are probably too simplifying to get a sufficiently accurate representation of the COVID-19 dynamic:

> Covid-19 is primarily transmitted through droplet infection, which indicates that the social contact structure in the population is important for the dynamics of infection. The compartmental model does not take into account variation in contacts between people, which occur in a society where few individuals could have many contacts and the majority have fewer contacts. This simplification in the model, i.e. a homogenous contact structure, usually results in a somewhat faster growth

> of an epidemic than if heterogeneity is included in the model. The model, therefore, runs the risk of overestimating the speed of the outbreak in the Stockholm region. This is not included in the specified confidence intervals, as a confidence interval cannot report such uncertainties (FoHM 2020b: 23).

Acknowledging that a model is overly simplifying to fully satisfy a certain purpose and yet choosing it over less simplifying alternatives indicates that the chosen model has other advantages that are more important for one's purpose and under prevailing conditions than the sacrifice in similarity. In the following, I will discuss five epistemic reasons that are all connected to simplicity considerations—thus arguing that FoHM's model choice indeed was motivated by a similarity-simplicity trade-off.

4.1 Lack of Reliable Data

The first reason for choosing compartmental models over agent-based ones is the lack of information needed for specifying some of the agent-based parameters. The latter contain a much larger number of parameters than the former for at least two reasons. First, agent-based models like *MicroPox* or *MicroSim* contain a lot of individual and institutional structures—e.g. demographic data and potential meeting places like dwellings, hospitals, offices, public transport—left implicit in compartmental models (Brouwers 2005, table 1). Second, these structures can take heterogeneous values for e.g. transmission probabilities at different places, individual probabilities of visiting the emergency when feeling ill or the propensity to travel (ibid.). This provides agent-based models with a much higher *potential* to represent social contact networks and individual heterogeneities. But this modelling potential also imposes high demands on measurement and data provision. Only if sufficiently reliable information is available can the model potential be actualized into a useful model:

> The [agent-based models'] usefulness would be huge if you had more knowledge about individual differences, like susceptibility, immunity, etc. But we don't have that. Hence no reason to use those models! (Brouwers IFFS talk 2020)

> The more information we have, the more possibilities we have, but we are still not at the point where we see that it is useful to switch to agent-based models because there is still so much that we do not know. (Brouwers interview 2021: 5).

At the beginning of the pandemic, the lack of data about the specifics of COVID-19 and its viral SARS-CoV-2 agent was particularly acute and led FoHM to choose models that did not explicitly require this information, instead replacing them with simple random-mixing assumptions. However, it is quite common for agent-based models to suffer from such data deficits, even outside of emergency situations. Take for example Brouwers' earlier *MicroPox* model:

> The data set contains […] no information about which school a child is enrolled in. Therefore, we must generate a proxy for this connection in the *MicroPox* model [We use] proxies for relations that are important to include in the social network, but for which we have no real data (Brouwers 2005: 73).

> Instead of specifying these free parameters with empirically well-founded information, they are filled through plausibility considerations that *might* be correct, although the modeler has little reason that they actually are (for further discussion of these proxies, see Grüne-Yanoff 2021).

For obvious reasons, it is more difficult to obtain reliable information of infection rates at different locations, instead of obtaining an average over the whole population, as used e.g. in the *SEIR* model. To explicitly include additional parameters that cannot be specified based on reliable evidence then constitutes a source of uncertainty and error that the FoHM modelers sought to avoid:

> what would be the use of using an agent-based model if you don't have that information? [One might] overestimate the risk that [agents] are part of large families or they work in certain places, and […] what you get out from the model would be based on that misassumption. So I would say: […] it's the risk of introducing more errors when using individual-based models, when you don't have additional information (Brouwers interview 2021: 10-11).

Model choice thus was motivated by weighing potential errors from different modeling strategies. Clearly, the FoHM modelers saw the simplifications of the compartmental models as a threat to relevant similarity between model and target, and thus as potential errors (as expressed in the quotation at the end of section 4). But the richer and more flexible structure of the agent-based models only offered the *potential* of building more similar models; this potential similarity could however only be realized with sufficiently good data (Grüne-Yanoff 2021). As such data—e.g. about individual differences in susceptibility and immunity or individuum-based social contact networks—was not available, the demands of high parameter specificity posed its own danger of generating errors. The modelers apparently consider the latter the graver threat, and thus chose simpler models over potentially more similar ones.

4.2 Avoiding Overfitting

The above problem concerns the unavailability of data for an independent determination of certain model parameters—e.g. the determination of the proportion of recorded and non-recorded cases based on the *Gloria* studies (section 3.2). However, many model parameters are not determined that way, but by estimation. For example, the early *SEIR* model estimated infectivity rate parameters θ, δ, ε and b_t from recorded infections and the measured proportion of recorded and unrecorded cases, using least-squares regression analysis (FoHM 2020b: 12-14).

For such model estimates, a too large number of parameters poses an additional source of error. It starts from the fact that data is to some degree always contaminated by random measurement error. Random error can in principle be reduced through increased sample size, but sample sizes are often limited and the size of random error is not known for many data-generating processes. A concern for the modeler therefore is to not *overfit* the model to the data set—i.e. to not fit the parameters in such a way that the model begins to describe the random error in the data rather than the relationships between variables. Such a result is more likely the higher the number of free parameters is in the model (Zucchini 2000). Overfitting has at least two negative consequences. First, an overfitted model describes the relationships between variables less accurately than an ideally fitted one. Second, overfitting fits a model to a specific set of data, including all its idiosyncrasies, thus deteriorating its abilities to predict future data. In both cases, increasing the number of parameters, and thus decreasing simplicity, makes the model sensitive to additional error (Grüne-Yanoff 2021).

Overfitting also illustrates that the trade-off in model choice is between simplicity and *potential* for similarity between model and target. A model that contains more parameters in principle can of course be fitted better to a (complex) target than a model with less parameters. But the practices of model estimation and calibration put a limit to how far this ideal can be reached. Not only is the data sample limited, it is also contaminated with noise. Therefore, the potential of a model with many parameters can rarely be fully realized. The modeler choosing a simpler model thus does not trade off genuine similarity, but rather only the (often unrealizable) promise of potential similarity (Zucchini 2000: 45).

Overfitting might not have been a concern for FoHM at the beginning of the pandemic. I could not find evidence that they specifically worried about random error in the data, nor whether they explicitly compared the number of free parameters in the *SEIR* and the agent-based models (it isn't even obvious how many of the latter's parameters would have to estimated). Yet the quality of the data, and the worry about parameter uncertainty played an important role, as I showed in the previous section; and these are considerations that also raise overfitting worries.

4.3 Easier Communicability

FoHM has well-specified client groups: parliament, the national government, other state authorities, regions, district councils and municipalities, district administrative councils and various interest groups (FoHM 2021c). FoHM not only provides these clients with facts, but also provides them with information about methods, so that clients trust the results and can explain to the public how they came about.

> [A]s soon as we publish results for one region, or a forecast or scenarios for a specific region, they will get questions from journalists, and they need to be able to answer those questions, and they need to understand what the model is showing and how it came about. So we have had lot of meetings [where we are] pedagogically going through how the modelling is performed, what data is used? What is a *SEIR* model? How has it been calibrated to real-world data? And what are the different scenarios? [...] they would be confident that what is in the report, the modeling, is something they kind of trust themselves. […] And they could say […] yeah, we have been part of this process (Brouwers interview 2021: 9).

To be able to achieve understanding and epistemic trust in their clients is a good reason to keep models simple. Simplicity here again concerns in the first place the number of parameters included in the model. The more parameters, the more computational steps are required to obtain a model result. With sufficiently many computational steps, a model can only be solved by a machine; and, in the more extreme cases, humans cannot even grasp how the machine arrives at a solution. Agent-based models, specifically those based on large data sets like *MicroSim*, are very much located at this extreme end. Such strong forms of *epistemic opacity* (Humphreys 2004) prevent the kind of shared understanding and epistemic trust that is part of FoHM's mission.

One might reply that ease of communication to clients, while important, does not concern the creation of knowledge and therefore does not constitute an epistemic value. The trade-off sketched in section 2, however, is about competing epistemic virtues of models; hence the above considerations should not count as relevant for that debate. Fair enough.

But ease of communication also concerns interaction within the knowledge-building team, and here it *does* constitute an epistemic value:

> being able to communicate within the group, but also with the managements at the FoHM, so we have a constant dialogue and they understand what we are doing in the model. Even Johan [Carlsson, the Managing director] understands, Anders [Tegnell, the state epidemiologist] understands what we are doing in the model: what are the drawbacks, what are the positive aspects of this model etcetera? […] So within the modeling group we have epidemiologists working tightly together with the modelers, like Jerker [Jonsson] for instance, who is an infectious disease doctor, who can directly advise on how we should implement the risk for hospitalization etcetera. What does it mean? How should we interpret what this data from Wuhan—what does it say? How can we think about this in terms of Sweden? […] So have a multidisciplinary working [group] together with the modelers, not only in the last stage but early, from the beginning (Brouwers interview 2021: 12-3).

The basic argument here is that the multidisciplinary team *as a whole* produces the relevant knowledge—neither modelers, computer scientists, epidemiologists or infectious disease doctors alone can produce it. This requires that each team member understands what the others are working with, and this crucially includes a basic understanding of the model. The more parameters and computational steps the model has, however, the less likely non-modelers will achieve this understanding, even if within-team communication is optimal. Thus, in a multidisciplinary team, there are good epistemic reasons to sacrifice some degree of (potential) similarity for the sake of understanding-facilitating simplicity.

4.4 Avoiding Fuzzy Modularity

There is a more specific kind of opacity, beyond that of generic parameter count, which arises in massive agent-based models like *MicroSim*. The more complex a model is, the more sub-components it has. In addition, when simulating a complex model, these model components run together and in parallel. Not each one, however, contributes to the model result independently. Rather, during a simulation, the components often exchange the results of intermediary calculations among one another—so that the contribution of each component to the model result is in turn affected by all the components that have interacted with it. Due to this interactivity, such agent-based models cannot be divided into separately manageable parts. Instead, these models represent a form of “fuzzy modularity” that makes understanding difficult (Lenhard and Winsberg 2010).

First of all, this is a problem for the explanatory power of agent-based models. Even if such a model could generate the explanandum quite accurately, it would be difficult to determine which of the modeled mechanisms contributed to the generated result. If understanding consists in identifying the mechanisms that created the explanandum, the fuzzy modularity of a model undermines improvements in understanding.

Now, policy makers perhaps do not have to worry about understanding. So why would fuzzy modularity be a problem for them? Due to it, users of agent-based models do not know how individual mechanisms contribute to the generation of a relevant effect. Knowing how individual mechanisms contribute is, however, of great importance for both (i) design and (ii) justification of interventions. First, without knowing how individual mechanisms contribute, the designer does not know where to intervene, because an intervention in a contributing cause can have several effects—through several mechanisms—that can reinforce or interrupt each other (Grüne-Yanoff 2021). Furthermore, they do not know whether the relationship between intervention and effect can be transferred to other contexts where some of the parallel mechanisms may work differently. This does not apply yet to FoHM's modeling, as they so far have modelled only few interventions.[6] But in later stages of modeling the epidemic and various interventions in it, where Brouwers saw some potential for employing *MicroSim* (section 3.2), fuzzy modularity might become an important argument against its use and for a continued sacrifice of (potential) similarity for the sake of simplicity.

4.5 Easier Error Detection

A final reason for trading off similarity for simplicity is that simplicity facilitates error avoidance and detection. All models used at FoHM today are computer-based; implementing models thus means programming them, and programming inevitably brings with it programming errors. Even though there might not be a correlation between complexity of code and number of bugs, there is a correlation with volume (e.g. "lines of code") and bugs (Fenton and Ohlsson 2000). Because simpler models have less code than more complex ones, on average simpler models contain less errors. Keeping computer-based models simple is thus a strategy for programming error avoidance.

Furthermore, once errors are in the code, it is easier to detect them in simpler models. The programming error detected at the April 21st FoHM press conference (see section 3.2) is a good example. By the time the press conference was over, five members of the public had already contacted Brouwers' collaborator to point out the same mistake. This required that the code was sufficiently simple so that educated laypeople could understand it and parse possible bugs. It would be difficult to imagine that something like this could happen with a massive agent-based model like *MicroSim*.

Some might reply that ease of error detection is not a primarily epistemic value, but rather only of pragmatic relevance. I disagree. Modelling tools' susceptibility to error, like that of any tool through which we hope to acquire knowledge, is of direct epistemic interest. Philosophers of science have accepted as much when they discuss strategies to avoid measurement error or ways of controlling background factors in experiments. They should treat strategies for avoiding programming error in the same vein. If indeed keeping one's model simple is an effective strategy in this regard, then modelers might well have a good epistemic reason for trading off some potential similarity for model simplicity.

[6] Exceptions are the introduction of vaccine compartments in some of the more recent scenario studies (FoHM 2020f, FoHM 2021a and 2021b) and Camitz' (unpublished) work on house quarantine.

5. Conclusion

The case study I presented in this paper illustrates that modeling methodology is a decision problem: modelers choose from a strategically prepared menu that model which they have reasons to believe will serve best their current purpose, under current conditions.

The examination of the menu developed at FoHM showed that these modelling alternatives differed mainly with respect to their simplicity—the number of parameters they contained—and therefore also with respect to the potential complexity of the target that they could represent. I argued that these differences are connected to the ongoing methodological discussion about whether modelers should trade off model-target similarity for the sake of increasing model simplicity—and thus about the validity of the KISS principle.

An analysis of the case study provided five reasons for choosing to engage in such a trade-off: lack of information, avoiding overfitting, avoiding fuzzy modularity, maintaining good communication, and allowing for error avoidance and detection. In addition, I argued for two observations: first, the purported trade-off is really between potential (not realized) model-target similarity; and second, each of these are indeed epistemic reasons. I conclude from these arguments that KISS, even in the time of COVID-19, is an epistemically important principle.[7]

References

Brouwers, L. 2005, “MicroPox: A Large-Scale and Spatially Explicit Microsimulation Model for Smallpox Planning”, in Ingalls, V. (ed.), *The Proceedings of the 15th International Conference on Health Sciences Simulation*, San Diego: Society for Modeling and Simulation International (SCS), 70-66.

Brouwers IFFS talk 2020, Personal notes from a talk by Lisa Brouwers, Institute for Future Studies, Stockholm, 10/9/2020 (unpublished material).

Brouwers interview 2021, Transcript of interview with Lisa Brouwers, Stockholm, 28/5/2021 (unpublished material).

Brouwers, L. and Liljeros, F. 2005, “The Functional Form of an Epidemic in a Real-World Contact Network”, Stockholm University working paper.

Brouwers, L., Camitz, M., Cakici, B., Mäkilä, K., and Saretok, P. 2009a, “MicroSim: modeling the Swedish population”, preprint arXiv: 0902.0901.

Brouwers, L., Cakici, B., Camitz, M., Tegnell, A., and Boman, M. 2009b, “Economic Consequences to Society of Pandemic H1N1 Influenza 2009-Preliminary Results for Sweden”, *Eurosurveillance*, 14, 37, 19333.

Burke D.S., Epstein J.M., Cummings D.A., Parker J.I., Cline K.C., Singa R.M., and Chakravarty S. 2006, “Individual-Based Computational Modeling of Smallpox Epidemic Control Strategies”, *Academic Emergency Medicine*, 13, 11, 1142-49.

[7] I thank Henrik Lundvall and Adam Lundström Ramirez for transcribing the interview, and Henrik Lundvall, together with two anonymous reviewers, for helpful comments on an earlier draft. I am most grateful to Lisa Brouwers, for her patience in answering my many questions, before, during and after the interview, and for her instructive comments on an earlier draft.

Chubb, M.C. and Jacobsen, K.H. 2010, "Mathematical Modeling and The Epidemiological Research Process", *European Journal of Epidemiology*, 25, 1, 13-19.

Edmonds, B. and Moss, S. 2004, "From KISS to KIDS—An 'Anti-Simplistic' modelling Approach", in Davidsson, P., Logan, B., and Takadama, K. (eds.), *Multi-Agent and Multi-Agent-Based Simulation: Joint Workshop MABS 2004*, Berlin-Heidelberg: Springer, 130-44.

Elliott, K.C. and McKaughan, D.J. 2014, "Nonepistemic Values and The Multiple Goals of Science", *Philosophy of Science*, 81, 1, 1-21.

Fasth, T., Ihlar, M., and Brouwers, L. 2010, "VirSim: A Model to Support Pandemic Policy Making", *PLoS currents*, 2.

Fenton, N.E. and Ohlsson, N. 2000, "Quantitative Analysis of Faults and Failures in a Complex Software System", *IEEE Transactions on Software Engineering*, 26, 8, 797-814.

FoHM. 2020a, "Skattning av behov av slutenvårdsplatser Covid-19", technical report, Public Health Agency of Sweden, release date 20/3/2020, revised 27/3/2020, https://www.folkhalsomyndigheten.se/contentassets/1887947af0524fd8b2c6fa71e0332a87/skattning-av-vardplatsbehov-folkhalsomyndigheten.pdf (last accessed 01/06/2021).

FoHM. 2020b, "Estimates of the Peak-Day and the Number of Infected Individuals during the COVID-19 Outbreak in the Stockholm Region, Sweden February—April 2020", technical report, Public Health Agency of Sweden. Artikelnummer 20055, revidering 1, release date 5/5/2020, https://www.folkhalsomyndigheten.se/contentassets/e1c3b83fa24f4d019e4842053ffd8300/estimates-peak-day-infected-during-covid-19-outbreak-stockholm-feb-apr-2020.pdf (last accessed 5/6/2021).

FoHM. 2020c, "Effekt av ökade kontakter och ökat resande i Sverige sommaren 2020", technical report, Public Health Agency of Sweden. Artikelnummer 21035, release date 15/6/2020, https://www.folkhalsomyndigheten.se/contentassets/29b815266baa4b409905c096be773df5/effekter-okade-kontakter-okat-resande-sverige-sommaren-2020.pdf (last accessed 5/6/2021).

FoHM. 2020d, "Estimates of the Number of Infected Individuals during the Covid-19 Outbreak in the Dalarna Region, Skåne Region, Stockholm Region, and Västra Götaland Region, Sweden", technical report, Public Health Agency of Sweden. Artikelnummer 200103, release date 1/7/2020, https://www.folkhalsomyndigheten.se/contentassets/e1702f53eea144cdb1ca2ef854b45c35/estimates-peak-day-infected-during-covid-19-outbreak-20103.pdf (last accessed 5/6/2021).

FoHM. 2020e, "Förekomsten av covid-19 i Sverige21-24 april och 25-28 maj 2020", technical report, Public Health Agency of Sweden. Artikelnummer 20105, release date 2/7/2020, https://www.folkhalsomyndigheten.se/contentassets/fb47e03453554372ba75ca3d3a6ba1e7/forekomstren-covid-19-sverige-21-24-april-25-28-maj-2020_2.pdf (last accessed 7/6/2021).

FoHM. 2020f, "Scenarier för fortsatt spridning—Delrapport 1", technical report, Public Health Agency of Sweden. Artikelnummer 20223, release date 21/12/2020, https://www.folkhalsomyndigheten.se/contentassets/fa087223be5c4bef8298ee2943f099ca/scenario-fortsatt-spridning.pdf (last accessed 5/6/2021).

FoHM. 2021a, "Scenarier för fortsatt spridning—Delrapport 2", technical report, Public Health Agency of Sweden. Artikelnummer 21035, release date 2/3/2021, https://www.folkhalsomyndigheten.se/contentassets/cc4a46bdfba54df38b9507c5d6242f77/scenarier-fortsatt-spridning-21035.pdf (last accessed 5/6/2021).

FoHM. 2021b, "Scenarier fӧr fortsatt spridning—Delrapport 3", technical report, Public Health Agency of Sweden. Artikelnummer 21087, release date 6/6/2021, https://www.folkhalsomyndigheten.se/contentassets/21443db264d84f7fbf495003 726b89f5/scenarier-for-fortsatt-spridning--delrapport-3.pdf (last accessed 5/6/2021).

FoHM. 2021c, "Brief Facts and Organization", https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/about-us/brief-facts-and-organization/ (last accessed 01/06/2021).

Giere, R.N. 1988, *Explaining Science: A Cognitive Approach*, Chicago: University of Chicago Press.

Goodman, N. 1972, "Seven Strictures on Similarity", in *Problems and Projects*, Indianapolis-New York: Bobbs-Merrill, 437-46.

Grüne-Yanoff, T. 2021, "Choosing the Right Model for Policy Decision-Making: The Case of Smallpox Epidemiology", *Synthese*, 198**,** 2463-84.

Humphreys, P. 2004, *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*, New York: Oxford University Press.

Hejblum, G., Setbon, M., Temime, L., Lesieur, S., and Valleron, A.J. 2011, "Modelers' Perception of Mathematical Modeling in Epidemiology: A Web-Based Survey", *PLoS One*, 6, 1, e16531.

Karlsson, E. 2020, "Chefen efter kaosdagarna: "Vi har kört nattarbete"", Aftonbladet, 23/4/2020, https://www.aftonbladet.se/nyheter/a/QobQ6J/chefen-efter-kaosd agarna-vi-har-kort-nattarbete (last accessed 01/06/2021).

Karlsten, E. 2020, "Johan Gieseckes kontrakt med Folkhälsomyndigheten—kan arbeta motsvarande en halvtid under 2020", https://emanuelkarlsten.se/johan-gieseckes-kontrakt-med-folkhalsomyndigheten-kan-arbeta-motsvarande-en-halvt id-under-2020/ (last accessed 01/06/2021).

Lenhard, J. and Winsberg, E. 2010, "Holism, Entrenchment, and the Future of Climate Model Pluralism", *Studies in History and Philosophy of Science*, *Part B: Studies in History and Philosophy of Modern Physics*, 41, 3, 253-62.

Levins, R. 1966, "The Strategy of Model Building in Population Biology", *American Scientist*, 54, 4, 421-31.

Matthewson, J. and Weisberg, M. 2009, "The Structure of Tradeoffs in Model Building", *Synthese*, 170, 1, 169-90.

Parker, W.S. 2015, "Getting (Even More) Serious About Similarity*", Biology and Philosophy*, 30, 2, 267-76.

Paola, C. and Leeder, M. 2011, "Simplicity versus Complexity", *Nature*, 469**,** 38-39.

Rich, B.R. 1995, *Clarence Leonard (Kelly) Johnson 1910-1990: A Biographical Memoir*, Washington, DC: National Academies Press, http://www.nasonline.org/publications/biographical-memoirs/memoir-pdfs/johnson-clarence.pdf

Rochefort-Maranda, G. 2016, "Simplicity and Model Selection", *European Journal for Philosophy of Science*, 6, 2, 261-79.

Weisberg, M. 2013, *Simulation and Similarity: Using Models to Understand the World*, Oxford-New York: Oxford University Press.

Zucchini, W. 2000, "An Introduction to Model Selection", *Journal of Mathematical Psychology*, 44, 1, 41-61.

Modeling Pandemic: Proximate and Ultimate Causes

Federico Boem

University of Florence

Abstract

In the understanding and prediction of a pandemic phenomenon, epidemiology is obviously the dedicated discipline. However, epidemiological models look at what we might call the proximate causes of the pandemic. On the other hand, the ultimate causes, those of an ecological, evolutionary, and socio-economic nature, are often too simplified or reduced to "minor" variables in epidemiological models. In this article, in dealing with a pandemic, we want to support the need to extend the study and design of responses to the ultimate causes and the disciplines that investigate them, with the hope of building an integrated approach for the future.

Keywords: Scientific modelling, Pandemic, Philosophy of medicine, Epidemiology, Ecology, Causation.

1. Introduction

The main goal of this article is to offer a different perspective on what are the possible *causes* of the COVID-19 pandemic. Generally (and in the first instance) the pandemic phenomenon has been approached as a medical/epidemiological problem. This is obviously understandable and also reasonable. In fact, this type of approach allows the scientific community and, in turn, policymakers to understand some salient aspects of the pandemic phenomenon that not only offer an epistemic advantage but are also essential to be able to think of strategies that aim to face and contain it.

From this point of view, it is therefore obvious that essential aspects are both the biological characteristics of the virus (such as its sequence) and the mechanisms and modalities of diffusion and infection. Specifically concerning phenomena of this type, scientific knowledge often focuses on the construction of *models*, both to explain and to predict such phenomena.

However this perspective, despite being visibly central and necessary, does not take into account those causal aspects of the pandemic that are more "distant" but must be seen as the context conditions that made the phenomenon possible in its actual realization. In this sense, specialists from fields other than medicine

Argumenta 6, 2 (2021): 57-77
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.boe

and epidemiology, such as theoretical ecologists, economists, and social scientists, have proposed to model the pandemic in the sense of trying to offer analysis for those factors that, even if of greater granularity, they are no less important or negligible.

The article is structured as follows. First, I will briefly present some established (to the scientific community) evidence about the nature of the SARS-CoV-2 virus and the pandemic. Secondly, I will describe what is generally meant by the activity of *modeling the pandemic*, especially from an epidemiological point of view. Next, I will introduce the distinction (originally developed by the naturalist Ernst Mayr) between "proximate causes" and "ultimate causes", and I will try to show how such a theoretical distinction can be useful in reference to the pandemic phenomenon. Fourth, I will present what I call the ultimate causes of the pandemic and what are the attempts at modeling them. Finally, I will try to show how these different aspects can contribute, not as alternatives but in a complementary way, in view of a broader, more complete, and adequate understanding of the pandemic.

2. Covid-19 Pandemic and SARS-CoV-2

On 11 March 2020, the World Health Organization (WHO) officially declared the Covid-19 pandemic. A pandemic is in fact a way of characterizing an epidemic that has specific characteristics. In general, the term, already etymologically, implies the tendency to spread everywhere and in a relatively short time. Therefore a pandemic occurs when some specific conditions are met. These are the presence of a highly virulent pathogen, the possibility of intra-specific transmission within the human species, and the lack of specific immunization towards the pathogen in the human population. Nevertheless, some experts point out that the term "pandemic" itself, although it may be still useful for communications in emergency situations, does not have a precise definition in quantitative and measurable terms (Singer et al. 2021). In this context, therefore, I will use the term "pandemic" in its broadest meaning (also according to the deployment of the WHO) of a global epidemic.

Faced with an emergency of this magnitude, alongside the studies that observe and try to describe the phenomenon, the other main activity of scientific research is to aim to figure it out. In other words, to understand its *causes*.

This, in turn, implies providing an explanation for the phenomenon but also building reliable predictions on its behavior. The two concepts, *explanation* and *prediction*, are obviously linked (intuitively, a well-founded explanatory model should also have good predictive power) but they must not be confused (Diéguez 2009, Douglas 2013, Findl and Suárez, 2021, Frigg and Hartmann 2020, Potochnik 2017, Shmueli 2010).

There are in fact, especially in disciplines such as computational biology, empirically predictive models but with little explanatory power. Conversely, explanatory models can be constructed that are not strictly predictive. Roughly speaking, within scientific practice, while explanatory models are those designed to test causal hypotheses about abstract constructs[1], predictive models just aim to forecast the behavior of a phenomenon (Potochnik 2017, Shmueli 2010). I will come back to these aspects further.

As a matter of fact, when investigating the causes of the pandemic, it is clear that epidemiology (among other disciplines) must be addressed.

Epidemiology is a discipline (which arises from the encounter of different areas of research) that studies the frequency with which certain pathologies occur in different groups of people and concerns the reasons for such scenarios. Based on these analyzes, epidemiology then builds models to plan and evaluate interventions in order to counter the spread of a certain disease or to prevent or treat it in those subjects in which it had developed[2].

Epidemiology obviously interfaces with other research areas, especially in the biological and medical sectors. In the case of an infectious disease such as Covid-19, one of the first steps is to understand the nature of the pathogen, its mechanisms of spread and infection, and its evolutionary origin.

During its development, it was learned how the Covid-19 was *caused* by a specific pathogen, which was then identified and classified as SARS-CoV-2. The SARS-CoV-2 coronavirus is a viral strain belonging to the subgenus *Sarbecovirus*, of the coronavirus subfamily (*Orthocoronavirinae*). Such a group is quite well known among researchers. In fact, several members of this set of viruses are responsible for various diseases (such as the common cold), including also quite serious diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) (Zhu et al. 2020, Sironi et al. 2020). In a short time, it was possible to determine the viral sequence of SARS-CoV-2, the main routes of diffusion, while there are still open hypotheses on its origin and on the steps regarding the transition to the human species (Sironi et al. 2020).

The surprise of Covid-19 must not suggest that pandemics are new phenomena. Pathologies of this type have accompanied the history of the human species since the Neolithic period. This period of human history is normally associated with the transition from a nomadic culture to forms of aggregation of a permanent nature. This is also the period in which the anthropogenic footprint on the environment has grown and the first forms of animal domestication are established. This step is essential given the zoonotic nature of Covid-19. Indeed, Covid-19 is a *zoonosis*, that is, an infection that originates in animals other than humans and is then transmitted to our species. This type of transmission can occur either directly (from species x to species Homo sapiens) or indirectly (through another intermediate species between the two). When this happens, we are in the presence of the phenomenon known as *spillover*. When a population of a given species, with its associated pathogens, comes into contact with a population of a different species, some pathogens of the starting species can adapt to a new species, generating a new form of the disease. Spillover is a fairly common occurrence in human history. In fact, over 60% of human viruses (including HIV and measles) are of zoonotic nature (Gibb et al. 2020).

[1] In saying this, I do not mean that the "causal view on scientific explanation" is the correct one. In making this distinction here I limit myself to describing a vision that is well represented within the scientific community (from the point of view of scientific practice) and thus not to take a position in the philosophical debate on this issue.

[2] See for instance https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/1-what-epidemiology (accessed April 27, 2021).

3. Modeling a Pandemic

As already stated, in order to *understand* a phenomenon (in the sense of being able to comprehend a part of its behavior in order to develop responses to it) such as a pandemic, alongside the biological characteristics of the pathogen and the knowledge on the functioning of certain biological mechanisms, scientists build *models*.

Without going into too much detail, a scientific model can be seen as some form of *representation* (for a more detailed discussion see Frigg and Hartmann 2020). In other words, models can be understood as forms of scientific representation that stand for a "portion of the phenomenic world", which is what one wants to represent [3]. Some scientific models are physical objects (enlarged or reduced) that represent the phenomenon under scientific investigation on a different scale. Other models instead try to capture properties of the object of scientific interest and to use analogous (even abstract ones) structures, often more manageable and manipulable, in order to act on the model and infer properties of the phenomenon or predict its behavior under certain conditions.

Accordingly, models must represent phenomena, but what does it mean that a representation is scientifically adequate? In fact, a good model does not always materialize by providing a faithful representation. Some models do not mimic the phenomenon to be represented but rather try to highlight certain properties (both to explain it and to provide predictions on its behavior). This is because, generally speaking, scientific modeling tends to display some kind of *idealization* (on this aspect see, among the others, Potochnik 2017). Philosophers have proposed and analyzed several types of idealization. As a matter of fact, for most of the philosophical debate, it is possible to refer to two main types of idealization: so-called the Galilean one and the Aristotelian one. Simplifying, the Galilean idealization implies a form of distortion in the analysis and the representation of the phenomenon (e.g. considering the spread of the virus in a uniform way over the entire population concerned). The Aristotelian idealization, on the other hand, involves building a model in which some relevant properties of the phenomenon are privileged, leaving out other real properties but considered not involved with its explanation or prediction (for example, understanding the rate of spread of the virus does not require detailed knowledge of its molecular structure).

Nevertheless, it is not necessary here to go into such details. Thus, by simplifying it, idealization in this context means that the model selectively represents certain properties of the object excluding others, or it operates distortions/simplifications.

Indeed, mathematical models, a kind of model largely used in epidemiology, are usually idealized models. Simplifying a bit, a mathematical model is a representation of a certain object, process, or phenomenon through a formal (and often quantitative) structure. There are obviously many ways to use mathematics to build a model, but in general, we can say that the construction of a mathematical model will start with the choice of some elements, considered fundamental, of the reference system and with the determination of the possible relationships between them.

[3] There is a debate whether all models should be seen as representations. For the scope of this paper there is no need to further develop this distinction. However, for a discussion of this aspect see Grüne-Yanoff 2013.

Thus, roughly speaking, in epidemiology a model is a mathematical construction trying to represent some parameters (considered relevant) involved in the genesis and subsequent development/behavior of the phenomenon studied, such as infectious diseases.

There are many different epidemiological models. One of the most used (also used to offer Western governments the first estimates on the behavior of the pandemic), is the so-called *compartment model*. Simplifying, this kind of model describes the progress of an epidemic on the basis of specific *assumptions* about the infection. Such assumptions, such as the mode of transmission or the infectious capacity of the virus, do depend on the *empirical data* collected. Thus, the stronger and more reliable the data, the more robust the assumptions will be. Subsequently, based on these assumptions, the population is divided into epidemiological groups or *compartments*. For an infection such as SARS-CoV-2, a standard model divides the population into 3 distinct groups: there are the *susceptible* (those who run the risk of becoming infected), the *infectious* (those who have already been infected and who can spread the virus), and *recovered* (which includes those who no longer transmit the virus, either because they recovered or because deceased). This standard epidemiological model is also called "SIR" (from susceptible, infectious, recovered)[4]. Normally, diffusion analysis is based on the first consideration that the risk of infection is an internal characteristic of the system. This means that the number of those who are infectious and those who can transmit the virus are the two initial factors to consider in determining the risk of infection. As a matter of fact, a model is a dynamic tool. Since the number of infected varies over time, it also affects the value given to the risk of infection. Obviously, this situation represents a simplified and idealized scenario. And it should be because it is a model. Indeed, in the real phenomenon, there are several other factors that influence the risk of infection. For example, specific health policy interventions, such as lockdowns, curfews, physical distancing, the obligation to wear masks, the prohibition of gatherings, etc., all have an impact on the progress of the epidemic. Surely, when building a model, not all the relevant factors could be known. Therefore, depending on the aim, a good model might need to be updated, to include some of these factors. However, the fewer parameters a model has, the more manageable and applicable it becomes. The choice of a model, therefore, depends on various factors and on trade-offs between different epistemic needs (e.g. applicability vs adequacy). Indeed, an effective model is obviously based on the collection of certain data. However, data alone do not say anything, since it is crucial to know where they come from and how to use them. Thus, it is also extremely important to find out where and how data have been originated, i.e. information on the collection strategy adopted for data is required. Next, data management usually implies certain formal tools (such as statistics). But statistical analysis cannot be simply applied out of the blue. Rather, it needs the choice of a model. In order to decide which model to use (a decision that might involve, as in this case, the need of a higher predictive capacity) it is fundamental to recall that every model is based on specific assumptions, degree of accuracy, and applicability. Assumptions are aspects given for granted that should serve as the empirical background. However, there will always be a tension, between those who are experts in the phenomenon (such as virologists and public

[4] https://nautil.us/issue/84/outbreak/whats-missing-in-pandemic-models (accessed May 2, 2021).

health scholars) and the modelers (such as theoretical epidemiologists, statisticians, etc.), on what aspects should be considered in the model and what elements can be neglected/reduced. Next, modelers themselves could disagree on the granularity and precision of their tools: i.e. the different values given to approximation. Finally, another source of the debate can come from discussions taking place when model outcomes become available, by considering the degree of applicability of the model (e.g. how much is similar to the target phenomenon or its manipulability in relation to its empirical adequacy) (on these aspects see, among the others, Frigg and Hartmann 2020, Potochnik 2017).

For example, a model with many parameters will need a massive amount of data and therefore will be more empirically supported, but perhaps it will be more difficult to build and less useful. Conversely, a model with few parameters will need fewer data to function and provide predictions, but its degree of distortive power will be higher and therefore it will be more difficult for it to provide robust indications. In this case (but it is not the only one) the difficulty of building effective (in terms of prediction) and accurate (in terms of empirical adequacy) models is given by the need (undoubtedly not easy), to harmonize constraints, methods, needs, and objectives of different disciplines (such as mathematics, virology, immunology, epidemiology, pharmacology, and medicine).

To understand how much a model is dependent on its assumptions, consider this case. In March 2020, the famous Imperial College model (the first to try to understand the Covid-19 epidemic) was produced by Neil Ferguson and his group (Adam 2020, Ferguson et al. 2020). According to this study, the Covid-19 epidemic, in the absence of specific containment measures, would have produced (in the following months)[5] around 510,000 deaths in the UK and more than 2 million in the US. This data also did not include the possible deaths resulting from the impact of the epidemic itself on the health system. Concerning the Italian situation, the model predicted more than 250,000 deaths (in total), if a strict lockdown had not been applied. The same model estimated, in the presence of quarantine, up to 30,000 deaths in a peak week with as many hospitalizations in intensive care (Ferguson et al. 2020).

Fortunately, this model turned out to be quite wrong in the predictions. However, this is not because the scientists worked improperly (or at least it is not just that), but because of the types of *assumptions* made. For instance, concerning the Italian case, the model assumed that children transmitted the infection exactly like adults. A fact that has proved false, but which was not known at the time of modeling and which was not so foolish to suppose. Furthermore, the model considered the Italian territory too homogeneously, treating high-density regions in the same way as the less populous ones. Finally, given the health fragmentation of the country (for which health policies and their organization are organized on a regional basis), the model did not consider the differences in response possibilities and resources between the different regions.

Indeed, a few months later Ferguson commented that the first model was being adapted from an earlier model used to simulate a flu pandemic. Given the need to generate a model in a short time, but being in the absence of specific data (such as the characteristics of the virus, etc.), it was necessary to build it starting from some *previous assumptions*, considered reliable and of a similar nature (e.g. a pathology which has many characteristics similar to Covid-19) (Chawla 2020).

[5] Roughly, from April 2020 to August 2020.

As a matter of fact, the problems concerning the construction and the application of the model show very well how real the potential risk of *scientific induction* is (but also how difficult it is sometimes not to take it).[6]

4. Epistemic Issues with Models: Causality

Regarding epidemiological models, the physician and philosopher Jonathan Fuller highlighted how the difficulty of modeling something like a pandemic also lies in some crucial *epistemic choices*. For example, the spread of SARS-CoV-2 obviously depends on the mechanisms of infection of the virus (and therefore on the interaction between these and human biology) but also on human behavior. Indeed, according to Fuller,

> Yet more sophisticated disease-behavior models can represent the behavioral dynamics of an outbreak by modeling the spread of opinions or the choices individuals make. Individual behaviors are influenced by the trajectory of the epidemic, which is in turn influenced by individual behaviors (Fuller 2020).[7]

This also means not only thinking deeply about the assumptions that are made and why they are made, but also trying to use different models to capture diverse aspects of the phenomenon. As Fuller recalls, alongside the compartment models, *multi-agent models* were also used during the pandemic, which tries to capture and represent the behavior of individual citizens (also in response to the different contexts in which they operate), and *curve-fitting models*, which on the basis of the trend of infections, considered similar in certain aspects, build possible scenarios on the current one.

Leaving aside other problems, the question I want to address here is whether these models offer any representation of *causal* links. In other words, whether these epidemiological models are *causal models*. Before doing this, certain theoretical premises should be discussed.

The concept of *cause* is as central as it is ancient in philosophical reflection. Simplifying, by "cause" it is generally meant something or a process that determines a certain effect. In other words, the cause would represent the origin or the condition of possibility of the occurrence of another fact. However, this conception (in the simple sense of elements, being either processes or objects, such as "A causes B"), although it captures aspects common to all causal accounts, remains too vague to be applied operationally. Therefore, looking at the differences between the various ways of understanding the notion of cause, it is quite evident that the literature on the topic is boundless. Leaving aside David Hume's famous (and still relevant) foundational critique of the notion of causality (linked to the assumptions on the regularity of nature and therefore connected to the problem of induction, see footnote 6), in the contemporary debate it is possible to distin-

[6] The problem of induction, briefly the question concerning the degree of certainty to be ascribed to the results obtained by inductive reasoning, is one of the central problems of the philosophy of science and scientific methodology. Obviously, this is not the place to examine this issue in general. For a more in-depth discussion see Henderson 2020, Henschen 2021.

[7] https://nautil.us/issue/84/outbreak/whats-missing-in-pandemic-models (accessed May 20, 2021).

guish at least five lines of research on causality: the *probabilistic* account, the *manipulative* one, the *mechanistic* one, the *counterfactual* one and finally that the *causal networks* (for a more detailed discussion on these aspects see Campaner 2011, 2012). From the point of view of scientific practice, as regards the construction of models, the employed idea of causality is not always made explicit.

Moreover, as far as diseases (and therefore epidemiology) are concerned, the concept of cause is not static, but also reflects a historical development. Thus, it is possible to briefly outline both the evolution of the idea of *causality* and how it is represented. By looking at the development of medicine as a modern discipline and especially considering epidemiology, the type of account generally adopted, more focused on the description of health determinants and risk factors rather than their underlying mechanistic understanding, has been progressively accused to display, concerning causation, a lack of adequacy (Campaner 2011). This is also due to the fact that, in the past, scientists and physicians were prone to reduce causal factors to simple, monadic, proximate, and detectable ones (such as the presence of a specific pathogen as in the Koch's postulates). Moreover, the single individual as such has been, traditionally, the main focus (both in terms of investigation and explanatory target) of epidemiology, this resulting in a diminished consideration of other determinants of health and disease. On the contrary, disciplinary advancements have instead promoted a more dedicated interest in groups and populations (especially in relation to infectious diseases) making epidemiology a central discipline for hygiene and public health policies (Campaner 2011).

Indeed diseases seem not to be entirely explainable assuming they are mainly determined by a single *factor.* Indeed, there are cases in which the cause of a condition, such as smallpox, is somehow simple since no smallpox can take place without the peculiar virus presence. However*,* advancements in clinical research have shown how the *causes* of a disease (in the plural) should be rather seen as a set of *sufficient conditions*, which generate favorable scenarios for the development of the disease.

Following this perspective, in 2005, Rothman and Greenland argued that the attribution of causality, in epidemiology, should not be conceived as the top-down formulation of criteria aimed at determining the presence of a certain effect, but as the "measurement of an effect" (Rothman and Greenland 2005). Roughly speaking, Rothman and Greenland claim that the origin of a disease can be traced back to a "sufficient causal complex" (pictured as a "pie"), which is represented by the composition of several constituent causal factors. Accordingly, if all those factors occur together, then the disease process initiates. This complex is then a necessary requirement for the disease.

Despite its success and adoption by many epidemiologists, this perspective on causality has been criticized by Vineis and Kriebel (2006), suggesting that the situation depicted by Rothman and Greenland is certainly a possible scenario but it is also too reductive. Indeed, it is usually the case that the association of several factors is more complex than "a single pie", meaning that there might be several different sets of causes capable of forming a "sufficient causal complex" for the same disease.

Furthermore, the situation is even more complex when attempting to distinguish between the causal dimension of the disease as a single occurrence in a given individual and the disease in its occurrence at the population level. Concerning this point, Vineis and Kriebel (2006) in fact argue that there is no doubt that, on a population level, certain phenomena, such as tobacco consumption, constitute a causal factor of certain forms of tumors (particularly lung cancer). However, it cannot be always stated that a particular case of this type of disease is necessarily attributable to smoking.

This is, for instance, the case of the famous “hallmarks” of cancer (Hanahan and Weinberg 2011). These hallmarks should be conceived as those factors which, both individually and in combination, can determine the onset of neoplastic pathologies. Indeed, these hallmarks are elements that raise the chance to develop such a condition at a population level, but it may certainly be the case that a single patient does not present most of them. This is also because, when dealing with the causes of a disease, it is extremely difficult to discriminate between variables that can serve as determinants or confounders of the causal pathway leading to the onset of a clinical condition (Vineis and Krieber 2006).

Similarly, Hill's famous criteria can be read in this light (Hill 1965). In summary, these criteria are *temporality* (the cause must precede the effect); *consistency* (the association between risk factor and disease must be confirmed in different contexts); the *strength of an association* (ie an association between a presumed determinant of disease and the disease itself can be more or less "strong"); *specificity* (the constancy with which a specific exposure produces a given disease, obviously, the more the biological response to the presumed cause is constant, the more likely it is that the latter is an actual cause); *biological plausibility* (i.e. the fact that the alleged cause is likely to be framed in the context of biological knowledge on the subject and on the pathogenesis). According to a recent study (Shimonovich, Pearce, Thomson et al. 2020) rather than conceiving these criteria as conditions of causality, it would be more appropriate to think of them as *aspects* that must be taken into consideration when talking about causes.

Simplifying, we could say that, from a methodological point of view, epidemiologists would consider an empirical relationship (between a disease determinant and a parameter of occurrence) as causal, when it persists even after verifying (in principle) all possible confounding effects. Discriminating real causal effects from confounding factors may not be an easy task. Among others, a particularly interesting modeling approach in dissecting causal aspects from confounding effects is the one based on *direct acyclic graphs* (DAGs). In those statistics models based on DAGs, the graph nodes are the possible elements in play, while the arrows represent causal effects. Those models can be useful in offering a better representation of the causal paths, on which to build a quantitative estimation/strength of the association.

Philosophically, the question here concerns the confrontation of different and alternative causal accounts. On the one hand, in fact, it is certainly possible to try to derive knowledge of a causal nature starting from statistical models. This can be done, for instance, by assuming some form of a theory of causality based on regularity and therefore probabilistically tractable (see, among the others, Hájek and Hitchcock 2016)

However, it is also important to point out that *biologically significant* phenomena do not always mean *statistically significant* phenomena. This means that it is not always possible to capture relevant biological relationships, of a causal nature, through purely statistical methods. In other words, it is very difficult to derive all relevant causal connections concerning a biological phenomenon, deriving them simply from models of a purely statistical nature. This is because statistical models, based on the analysis of variance, have no direct way of discriminating (even qualitatively) the nature of the diversity of biological interactions. Indeed, considering the complex relationship between biological objects and their interaction with the surrounding environment (a bio-ecological relationship for which organisms shape the environment and are in turn modified by it), Vineis and Kriebel directly report:

> [A]nalysis of variance will correctly correspond to an "analysis of causes" (i.e., quantifying the relative importance of the main effects of genes, environment and their interactions) only when: (a) environmental exposure-response relationships are linear for individuals with each of the different genetic polymorphisms, and (b) the study includes a sufficiently broad range of exposures to provide statistical power to detect an interaction (Vineis and Kriebel 2006: 5).

The study of biological and ecological interactions, as determining factors in the onset and development of a disease, is therefore crucial for a causal investigation in epidemiology. On the one hand, the development of models capable of capturing the diversity of possible interactions is certainly central (probably different models will be more suitable for certain interactions than others). However, as Vineis and Kriebel (2006) recall, it is good to remember that interactions cannot be totally captured and consequently modeled by an approach that reduces them to elements that can be manipulated with statistics. Even the theoretical modeling in epidemiology, although it must be idealized and abstract for the explanatory and predictive purpose, cannot ignore a deeper knowledge of what are the biological and ecological notions of causality, concerning the phenomena in progress.

Finally, as Fuller recalls (2021), in the case of compartment models, such as those adopted to model Covid-19, there is also *epistemic friction* involving a clash between diverse accounts of causation. According to Fuller, this friction occurs because, on the one hand, those models are conceived as *causal* (by virtue of their formally representing the mechanism of infection and spread of the virus), and on the other hand not all scholars would admit that simple manipulation of the parameters of such models allows making *causal inferences* (with which to discover or determine new "causes" (previously unknown) regarding the pandemic). This is because compartment models can provide causal explanations when the underlying mechanism (which they are based upon) embeds a form of interventionist/manipulative causal account. Thus, playing with the "gears of the mechanism", such as the adoption of a particular policy as physical distancing or the use of masks, it is possible to evaluate their causal role in terms of produced effects. However, as Fuller recalls:

> [T]hese estimations simply involve manipulating model parameters and comparing what falls out of the model under different values, and 'causal inference' is typically thought to combine causal information with non-causal information to infer a novel causal conclusion. Thus, the idea that compartment models are causal models may be in tension with the idea that on their own they can do causal inference. If they are purely causal models, then we may intuitively think that we cannot infer new causal knowledge simply by manipulating them; any causal conclusions we derive must in a sense already be contained within the model. While we can hang on to the commitment that compartment models are causal models by accepting that manipulating parameters generates causal predictions and retrodictions rather than so-called causal inferences, it may be difficult to shake

> the intuition that we learn about novel causal relationships (including their quantitative strength) by tweaking model parameters (Fuller 2021: 47).

In researching the causes of the pandemic, therefore, one requires, as Fuller also suggests,[8] the need to think philosophically. This definitely means to adopt a general, critical attitude towards data and methods, but it may also mean taking a step back, and asking what kind of question is that which concerns the causes of a biological phenomenon (and thus also reflecting on possible different accounts of causation). Such a "mode of thinking" also entails (probably) going beyond the confines of epidemiology as such. However, it is to be hoped that these aspects can then be adequately included in epidemiological analysis.

5. Ernst Mayr's Revisited: Ultimate and Proximate Causes of a Pandemic

In 1961, the famous naturalist Ernst Mayr published an article (which later became a classic) on the concept of cause in the life sciences. Mayr proposes the idea that there are essentially *two types of cause* in biology. To better put it, he argues that there are two epistemic accounts of causal investigation in biology, irreducible to each other, both fundamental and, in a sense, complementary.

Following Mayr's terminology, a cause can be "proximate" or "ultimate". The proximate causes answer the question about *how* a particular phenomenon occurs. Mayr seemed to have in mind that proximate causes capture a sort of mechanistic causal link[9] precisely because they deal with mechanistic representations that allow scientists to unravel how certain phenomena take place) (Mayr 1961). For example, a proximate cause of SARS-CoV-2 infection is to be found in the "spike" protein that allows the virus to "enter" a certain type of cell. Another proximate cause might concern the mechanisms of diffusion through small particles of liquid contained in breathing or in a sneeze.

On the other hand, we have ultimate causes. It answers the question about the *why* of a given phenomenon. In this perspective, the ultimate causes, therefore, aim to explain the pandemic not in its etiological-epidemiological mechanisms but rather in the reasons/conditions that allowed such a global epidemic to take hold. According to Mayr, the ultimate causes are often the evolutionary causes of a certain biological phenomenon.

The exquisitely epistemological dimension of Mayr's account is evident. Indeed, the naturalist does not place the understanding of a biological phenomenon as a choice between these two alternatives. Rather, he argues that if the causes of a certain biological phenomenon are to be understood, it is essential to recognize that different causal links answer different questions, distinct but complementary, and that also answer different research methodologies. The composition of these perspectives would allow us to offer an understanding of the phenomenon in its complexity.

The distinction made by Mayr has greatly shaped the epistemic attitude of biologists from the second half of the twentieth century onwards. For example,

[8] https://nautil.us/issue/84/outbreak/whats-missing-in-pandemic-models (accessed May 2, 2021).

[9] In the sense of the physiological mechanisms that "govern the responses of the individual (and his organs)" (Mayr 1961: 1503).

even today, many scholars see molecular biology as a discipline that investigates proximate causes while evolutionary biology investigates the latter. The distinction is both extensive and much debated. While some recognize that it still captures a fundamental insight into causal aspects in biology, others argue that too rigid a reception may even hinder the development of research (Laland et al. 2011). It is also worth remembering that some of Mayr's concerns and observations also depend on the state of research in the 1960s. For example, as we saw in the previous section, more refined models (such as DAGs) allow us to easily represent situations with numerous causal factors. Particularly in epidemiology (which Mayr was not an expert of), as we have seen, the diversity of approaches and methods made possible a refined modeling, able to grasp and manipulate otherwise intractable relationships.

However, it is a fact that the epidemiological models adopted, although they have tried to include more and more variables of a socio-behavioral nature, have not, in Mayr's terminology, properly investigated *why* Covid-19 has become a global threat. In other words, the evolutionary and ecological aspects of the pandemic, although not neglected, did not constitute precise variables in the modeling.

6. Why Covid-19?

The ecological and evolutionary perspective on the pandemic seems to somehow confirm Mayr's proposal. If it is obvious that to act against the spread of the virus it is appropriate to work on the proximate causes of Covid-19, the possibility of preventing such a threat from happening again, and with this magnitude, lies in understanding the root causes. In other words, acting on proximate causes means on the one hand building models that allow the development of specific interventions, represented by specific variables, which can change the course of the infection, and on the other hand, working on the production of drugs and vaccines that defeat the virus itself in infected people and reduce the possibility of new infections. However, these approaches necessarily neglect the ultimate causes of the pandemic. In other words, to prevent a new pathogen from having such a great impact, it is necessary to pay attention to both the evolutionary dimension of viruses and their ecological dimension. Furthermore, this also involves trespassing into other disciplines. The economic production system, especially in Western countries (with the associated lifestyles) has been severely tested, as well as the organization of health systems and the very idea of public health policies.

In fact, according to a report by the United Nations Environment Program (UNEP) and the International Livestock Research Institute (ILRI), addressing just the proximate causes means, in fact, treating the health and economic *symptoms* of the pandemic but not its *causes* (UNEP and IRLI report 2020).[10] Instead, the causes are to be found in the disruption of ecosystems and the impact of human species on the environment. Limiting ourselves to containing the virus, mitigating its effects, or even eliminating it without having to deal with the organization of economic framework, public health policies, and without heavy interventions on the environmental contexts that create the conditions for *spillover*, we will soon find ourselves faced with other pandemics. The report states that the number

[10] https://www.unep.org/resources/report/preventing-future-zoonotic-disease-outbreaks-protecting-environment-animals-and (accessed April 28, 2021).

of "zoonotic" epidemics is generally increasing worldwide. Several new pathogens cause 2 million victims every year, mainly in the poorest countries. However, precisely for not having investigated the ultimate causes of the pandemic (and of other pandemics), the (Western) world found itself unprepared for the latest pandemic. Accordingly, Covid-19 has been treated as a purely medical problem, with repercussions on the economy and on people's lives but not as an ecological issue. According to ecologists, however, its origins are in the environment, in global food systems, and in the interactions between non-human and human-animal species (Gibb 2020). In particular, ecologists argue that the global expansion of agricultural and urban land (a phenomenon still growing and predominant in low-income countries) is one of the main reasons for the creation of zoonotic pandemic reservoirs, in which wild and domesticated species they are in close contact with each other and with human beings (Gibb 2020). Thus, a number of diverse, interconnected, factors form a causal web that is not easy to treat in a single way. As a matter of fact, overpopulation, deforestation, land consumption, the increase in urbanized areas and human intrusion into natural habitats, deforestation in favor of agriculture and intensive farming and mining are leading to the impoverishment of ecosystems and, in turn, fostering the conditions for the spread of pathogens.

Various researches in the field of ecology also show that the progressive alteration of global ecosystems is the main risk factor for the development of pandemics. In fact, the destruction of ecosystems very often involves the reduction (even the extinction) of some species, especially the more specialized ones. On the other hand, this involves the proliferation of more adaptable species which are more frequently the natural reservoirs of pathogens. According to this perspective, the conservation of biodiversity (with specific interventions, such as policies that limit or prevent deforestation and indiscriminate soil consumption), becomes a measure that acts directly on *ultimate causal factors*, significantly reducing the risk of future pandemics (Gibb 2020, Murányi and Varga 2021, Finlay et al. 2021).

According to a world program of the WHO, called "One Health", the future of research also in the medical field (especially with repercussions on public health) concerns the ecological aspect of diseases.[11] According to this perspective, the very concept of *human health* must be updated and integrated with *animal health*, and more generally with an ecological perspective that includes the "health of the ecosystem". It follows that pandemics must be tackled with a multidisciplinary strategy, keeping together epidemiology, climate sciences, species protection, and risk communication (Fronteira et al. 2021). This is particularly crucial considering that of the emerging pathogens, about 75% are of zoonotic origin. Furthermore, zoonotic pathogens are twice as likely to generate emerging diseases compared to non-zoonotic pathogens (Taylor et al. 2001).

Another aspect, distinct but obviously connected, concerns the more properly evolutionary dimension. Biological entities such as viruses are in fact almost ubiquitous in nature and interact with every known form of life. Furthermore, it is now established that viruses play a crucial role in the dynamics concerning the genesis and development of all living forms (Harris and Hill 2021), and some scholars have even suggested that they constitute one of the determining factors of the evolutionary process (Koonin and Dolja 2013). This allows us to make some considerations on the character of epidemics and pandemics. As

[11] https://www.who.int/news-room/q-a-detail/one-health (accessed May 10, 2021).

already mentioned, they are nothing new in the history of the human species. Furthermore, infectious diseases have contributed to determining the development of the human species (operating as a selective filter) but have also conditioned human nature itself by virtue of the biological possibilities of interaction between the human species with other living forms (Gilbert et al. 2012, Harris and Hill 2021, Brett et al. 2021).

Furthermore, according to many experts, relationships between human urbanization, public health, and biodiversity need also to be investigated. Aspects of an ecological nature are therefore intertwined with issues of a social and cultural nature, making it even more difficult to deal with this level of causality treatable by a single discipline. First of all, the social and technological modality with which the human species has configured itself (especially the following industrialization) establishes a situation that is particularly suitable for the spread of pandemics. Contemporary human societies are made up of millions (sometimes billions) of individuals concentrated above all in certain areas where they live in close contact and according to dynamics that provide for strong social and physical interaction (Brett et al. 2021).

No less important is the aspect concerning the organization of health systems. The pandemic has clearly shown how the model (especially Western) built more and more around poles of excellence, but not very attentive to the medicine and health of the territory and creator of situations of health inequality and inequality, was one of the causes of the global disaster. From this perspective, Covid-19 was more of a crisis in the organization of health systems than a medical crisis (El Bcheraoui et al. 2020, Pescaroli et al. 2021).

Furthermore, this issue affects not only the practical dimension but also involves some of the very foundations of public health (again, especially in the Western world), such as the concept of “hygiene” (Brett et al. 2021). If it is definitely true, in fact, that the development of public health (and the very promotion of the concept of "hygiene" have eradicated many infectious diseases and significantly increased people's life expectancy, it is equally true that this model, typical of a society progressively urbanized, consisting of increasingly mediated interactions, has also produced a significant decrease in the microbial ecosystem with which the human species has evolved, including a reduction in the biodiversity of the human microbiota (a phenomenon often associated with the onset of various diseases, above all autoimmune in nature, but also susceptibility to some infectious diseases). This theoretical framework proposes that the changes, over time, that different human populations have undergone, have led to a consistent loss of biodiversity of microorganisms. These changes concern some characteristic elements of contemporary life (especially in Western countries): urbanization, the indiscriminate use of antibiotics, the hygiene of living and working environments, the homologation of food (towards a greater presence of food industrially produced), and excessive consumption of alcohol and tobacco (Brett et al. 2021).

Finally, part of the ultimate causes of the pandemic is to be found in the purely social, political, and economic dimensions. Regarding this point, it is decisive to make some specifications in order to avoid simplifications or striking statements that could be supported by little evidence. To argue that aspects far from biological and epidemiological mechanisms (such as the ecological dimension) play an ultimate causal role on pandemic means (following the spirit of Mayr's theoretical distinction) to investigate those causal factors that have determined the possibility of a certain state of affairs rather than another. In other words, if it is obvious that the pandemic has its necessary and proximate cause in the Sars-CoV-2 virus, that pathogen has nevertheless been able to be as such and to generate a phenomenon such as a global pandemic due to a set of causes at a systemic level.

Therefore, this is why many experts claim that it is undeniable that the pandemic was, in addition to a health emergency, also a political and social emergency (Morens and Fauci 2020, Brett et al. 2021, Leach et al. 2021). According to some economists and human development scholars, this also entails the need to think differently about the growth modalities of human societies and what interventions are fundamental for a rethinking that acts precisely on the ultimate causes of the pandemic (Leach et al. 2021). In other words, to identify the causes of the pandemic, to develop tools to manipulate its effects and stem its origins, it is essential to act on some key nodes of the organization of society itself (especially in the West).

For some experts, this means above all recognizing the limits and inconsistencies of the current economic growth model (also due to the aforementioned repercussions on the environment and biodiversity). This also means promoting forms of greater awareness (citizen empowerment) and participation by citizens in public policies, especially health. As some scholars write:

> Where traditional approaches to development have been top-down, rigid and geared towards narrow economic goals, post-COVID-19 development must be centered on a radically transformative, egalitarian and inclusive knowledge and policy (Leach et al. 2021: 1).

Indeed, as numerous specialists have noted in recent months, the impact of the pandemic has not been the same for all individuals, nor for all affected countries. The disparities and inequalities (both social and economic) of the different contexts have created a scenario that is anything but homogeneous. Therefore, many have argued that Covid-19 should not be considered a pandemic but rather a *syndemic* (Bambra et al. 2021, Fronteira et al. 2021, Griffith 2021, Horton 2021, Islam et al. 2021, McMahon 2021).

The notion of syndemic was originally developed by the medical anthropologist Merrill Singer in the 1990s. More recently, he and colleagues have proposed a model of approach to syndemic diseases (Singer et al. 2017). This means scientifically examining the *biosocial complex*, formed by the interaction of pathologies with social and environmental factors (either parallel to the onset of the disease or resulting from it). According to this perspective, this implies a new and different conception of the disease itself, not as a process sharply distinct from others, but rather a frame that puts it in relation to the other pathologies and the social, political, and economic contexts in which the disease occurs. The term "syndemics" therefore wants to emphasize the synergistic effects with which these different factors combine and their consequences. This translates into the study of why some pathologies focus on particular individuals or groups, and the ways in which contexts in which social inequality and economic disparity are determining factors in estimating the incidence of the disease.

This perspective obviously proposes to manage health emergencies in a different way (also from a causal point of view). As Singer and colleagues write:

> A syndemic approach provides a very different orientation to clinical medicine and public health by showing how an integrated approach to understanding and

> treating diseases can be far more successful than simply controlling epidemic disease or treating individual patients (Singer et al. 2017: 947).

7. Other Models and Possible Integrations

The factors that refer to what we have defined as the *ultimate cause*s of the pandemic appear crucial not only to understanding the pandemic itself but also to its management and to averting possible new future pandemics. In fact, as some scholars suggest (Leach et al. 2021), the understanding of the pandemic as a complex and global phenomenon (not only in its epidemiological meaning), requires a more in-depth study on the structural dynamics that concern various aspects (connected to each other) such as internal human interactions (e.g. social and economic relations) and those towards the biological world in an ecosystemic perspective (i.e. looking at human activity both as the cause of certain phenomena and as shaped by those phenomena themselves). Taking into consideration not only Covid-19 but also previous experiences (such as Sars and Zika), this also means recognizing that the elements of ecological disturbance (such as the reduction or destruction of ecosystems) are closely linked to material constraints, choices concerning policies, and economic, political and social conditions (Zabaniotou 2020, Leach et al. 2021). In fact, a response to the pandemic that deals only with its proximate causes, without questioning the global model that generated it, unequivocally linked to the kind of factors already mentioned, is not only incomplete but is limited to dealing only with the visible and symptomatic aspects, thus completely neglecting the triggering elements of the phenomenon.

From the point of view of scientific understanding, this also means that the modeling of a single level of description of the phenomenon, such as the epidemiological one, does not appear sufficient. As some researchers have pointed out (Engen et al. 2021), the compartment models used in epidemiology do not seem suitable to adequately represent the phenomenon. This is not to be understood in a simplistic way, as if it were a theoretical oversight or methodological neglect. On the contrary, this feature reflects the fact that the variables, parameters, and conditions are too many to be included in the model in a consistent way, also considering the necessary distortions of the model and the scarcity of more in-depth information (and not always available) on aspects necessary for the operation of the model itself. As a matter of fact, theoretical ecology, which daily works with the challenge of dealing with complexity without making excessive reductions, has over time developed models and approaches to contemplate the so-called "noise", which is one of the main characteristics of complex systems and in particular of biological ones. This implies to switch from more fixed and deterministic approaches to stochastic, noise-inclusive modelling strategies. Thus, according to some scholars in this field, applying this type of modeling to epidemiology can provide valuable tools both for measuring how an epidemic can take hold and for providing predictions about its development and possible responses to it (Engen et al. 2021).

On the other hand, there is no need to load a disciplinary perspective with miraculous or salvific properties. Indeed, there is also a debate within theoretical ecology on the explanatory and predictive scope of models (Schuwirth et al. 2019). As with epidemiological models, ecological models also make assumptions and approximations and cannot be considered comprehensive solutions, especially if not discussed with other experts coming from field studies and from experimental research. Moreover, not all ecological models are the same. For instance, recently there has been a quite intense debate concerning the feasibility and the pertinence of *species distribution models* (SDMs) for unravelling crucial features of Covid-19 (Araújo et al. 2020, Carlson et al. 2020a, 2020b). SDMs are tools to model complex "objects" such as habitats that have been progressively adopted in biomedical geography, linking ecological determinants to cases of epidemiological interest. Despite their success, some scholars have recommended attention to their adoption in the case of covid, given the scarcity of information still available to be able to model more substantially some crucial aspects of transmission (Carlson et al. 2020a). On the other hand, other researchers (Araújo et al. 2020), while recognizing the limitations and dangers of generalizations coming from ecological models (often less able to provide a mechanistic understanding of certain phenomena, as is the case with epidemiological models), have argued for the need to recognize the specific epistemic virtues of these tools and to consider the results of both approaches in a more open and interdisciplinary way. In this sense, in defending a pluralistic approach, which contemplates the use of different types of models (also from different disciplinary sectors) these authors write:

> While correlative models can provide insight concerning the environmental persistence of the pathogen (thus affecting spread of the disease), mechanistic approaches allow projecting numbers of infections and fatalities as a function of management policies. Rather than building walls across scientific disciplines, building bridges will be more effective to understand the spread of SARS-CoV-2 and its effects on human health (Araújo et al. 2020: 1153).

Accordingly, the adoption of ecological models (also in their diversity) should therefore not be conceived in contrast with the epidemiological description as much as in an integrative sense. For example, some ecological models (Coro 2020) can provide additional information (for instance on certain environmental aspects and biogeographical conditions that could favor or limit the spread of the virus) that is not traditionally contemplated in epidemiological modeling.

However, the ecological perspective (strictly speaking) is not the only relevant dimension for future modeling. Indeed, in order to build pandemic management policies (both at an emergency level and of a more structural layer), ecosystemic factors are intertwined with social and economic ones.

Regarding this point, it seems obvious that a single modeling that integrates all these variables in a complete way is almost impossible (and even where it was feasible, it would be difficult to use.). Therefore, public health specialists are starting to develop overarching operational schemes that can serve as a meta-theoretical framework capable of considering heterogeneous data together with a view to their consistent and coordinated use (Raboisson and Lhermie 2020).

In conclusion, as Campaner (2011) also recalls, the objective of integrating epidemiological and ecological accounts is to provide an integrated perspective of the scientific explanation that contemplates different levels of description without imposing forms of reductionism.

The complexity of a phenomenon such as a pandemic in fact implies that its explanation (in the sense of exhibiting a causal structure) cannot only concern those obvious and *proximate* causal aspects, but also those that make a difference in the way the pandemic itself manifests and develops (i.e. the *ultimate* causes).

This aspect also helps to bring out the first and foremost epistemological (rather than ontological) perspective of modeling and the need for integration at this level (rather, here too, than at the ontological one). In fact, as Campaner (2011) always reminds us, if it is certainly true that the entire causal dynamic of the pandemic is entirely given and objectively traceable (as a theoretical possibility), the explanation of its aspects will also be dictated by the context, interests and disciplinary approaches, with their differences.

8. Conclusion

The Covid-19 pandemic has proved to be an unprecedented global threat. Both in health terms and in terms of impact on human life and its organization, on a global scale.

The scientific community has tried to understand this partly totally new phenomenon by designing experiments to dissect its properties (such as the virus sequences or its mechanism of infection) and building models to understand its general behavior. These efforts, taken together, have tried both to *explain* the pandemic and to *predict* its progress, with the idea that these two concepts are the key for addressing a phenomenon, and for developing the capacities to control it.

Faced with an infectious threat of this magnitude, epidemiology has obviously been one of the essential resources to manage the emergency. Through the construction of various types of models it has been possible to try to evaluate the effectiveness of certain measures, the ineffectiveness of others and to plan not only health policies but all the activities that govern modern societies.

In their attempt to capture the causes of the pandemic, epidemiological models, although indispensable, work on those that explain the "how" of the occurrence of Covid-19, but neglect the study of the reasons behind Covid-19.

In this article, I have tried to outline how the fact of the pandemic also requires an attempt to answer the "why" of its being. In other words, an attempt to incorporate, within the scientific explanation, also the ultimate causes of the phenomenon: namely the ecological, evolutionary, and socio-economic factors related to the pandemic. This effort entails an ecosystemic perspective in at least two meanings. On the one hand, this perspective is purely disciplinary, that is, it concerns the methods and approaches used by the sciences of complex phenomena, such as ecology and by the social sciences. On the other hand, the reference to the systemic character refers to the epistemological level, that is to conceive the scientific explanation as organized on several levels, both in terms of granularity and of shape according to specific interests (also in relation to their ability to offer manipulations of the phenomenon in question).

This is not just a theoretical wish or a declaration of intent. It must be translated into effective practices. Thus, such an effort will require not only interdisciplinarity as such but also new integration strategies, concerning both differences in data production methods and diversities in methodological approaches.

References

Adam, D. 2020, "Modeling the Pandemic: The Simulations Driving the World's Response to COVID-19", *Nature*, 580, 316-18.

Araújo, M.B., Mestre, F. and Naimi, B. 2020, "Ecological and Epidemiological Models Are both Useful for SARS-CoV-2", *Nature Ecology and Evolution*, 4, 1153-54, doi: 10.1038/s41559-020-1246-y.

Bambra, C., Riordan, R., Ford, J., and Matthews, F. 2020, "The COVID-19 Pandemic and Health Inequalities", *Journal of Epidemiology and Community Health*, 74, 964-68, doi: 10.1136/jech-2020-214401.

Brett F.B., Amato, K.R., Azad, M., Blaser, M.J., Bosch, T.C.G., Chu, H., et al. 2021, "The Hygiene Hypothesis, the COVID Pandemic, and Consequences for the Human Microbiome", *Proceedings of the National Academy of Sciences of the United States of America*, 118, 1-9, doi: 10.1073/pnas.2010217118.

Campaner, R. 2011, "Causality and Explanation: Issues from Epidemiology", in Dieks, D., Gonzalo, W., Uebel, T., Hartmann, S., and Weber, M. (eds.), *Explanation, Prediction, and Confirmation*, Dordrecht: Springer.

Campaner, R. 2012*, La causalità tra filosofia e scienza*, Bologna: CLUEB.

Carlson, C.J., Chipperfield, J.D., Benito, B.M., Telford, R.J., and O'Hara, R.B. 2020a, "Species Distribution Models are Inappropriate for COVID-19", *Nature Ecology and Evolution*, 4, 770-71, doi: 10.1038/s41559-020-1212-8.

Carlson, C.J., Chipperfield, J.D., Benito, B.M., Telford, R.J., and O'Hara, R.B. 2020b, "Don't Gamble the COVID-19 Response on Ecological Hypotheses", *Nature Ecology and Evolution*, 4, 1155, doi: 10.1038/s41559-020-1279-2.

Coro, G. 2020, "A Global-Scale Ecological Niche Model to Predict SARS-CoV-2 Coronavirus Infection Rate", *Ecological Modelling*, 431, 109187, doi: 10.1016/j.ecol model.2020.109187.

Diéguez, A. 2013, "When Do Models Provide Genuine Understanding, and Why Does it Matter?", *History and Philosophy of the Life Sciences*, 35, 599-620.

Douglas, H.E. 2009, "Reintroducing Prediction to Explanation", *Philosophy of Science*, 76, 444-63, doi: 10.1086/648111.

El Bcheraoui, C., Weishaar, H., Pozo-Martin, F. and Hanefeld, J. 2020, "Assessing COVID-19 through the Lens of Health Systems' Preparedness: Time for a Change", *Globalization and Health*, 16, 1-5, doi: 10.1186/s12992-020-00645-5.

Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., et al. 2020, "Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand", *Imperial College COVID-19 Response Team*, March, 1-20.

Frigg, R. and Hartmann, S. 2020, "Models in Science", in Zalta, E.N. (ed.), *The Stanford encyclopedia of philosophy (Spring 2020)*, https://plato.stanford.edu/archives/spr2020/entries/models-science/.

Fronteira, I., Sidat, M., Magalhães, J.P., de Barros, F.P.C., Delgado, A.P., Correia, T. et al. 2021, "The SARS-CoV-2 Pandemic: A Syndemic Perspective", *One Health*, 12, doi: 10.1016/j.onehlt.2021.100228.

Fuller, J. 2020, "What's Missing in Pandemic Models", *Nautilus*, 6, http://nautil.us/issue/84/outbreak/whats-missing-in-pandemic-models.

Fuller, J. 2021, "What Are the COVID-19 Models Modeling (Philosophically Speaking)?", *History and Philosophy of the Life Sciences*, 43, 1-5, doi: 10.1007/s40656-021-00407-5.

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., and Rinaldo, A. 2020, "Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of

Emergency Containment Measures", *Proceedings of the National Academy of Sciences of the United States of America*, 117, 10484-10491, doi: 10.1073/pnas.2004978117.

Gibb, R., Redding, D.W., Chin, K.Q., Donnelly, C.A., Blackburn, T.M., Newbold, T., and Jones, K.E. 2020, "Zoonotic Host Diversity Increases in Human-Dominated Ecosystems", *Nature*, 584, 398-402, doi: 10.1038/s41586-020-2562-8.

Grüne-Yanoff, T. 2013, "Appraising Models Nonrepresentationally", *Philosophy of Science*, 80, 850-61, doi: 10.1086/673893.

Hájek, A. and Hitchcock, C. 2016, *The Oxford Handbook of Probability and Philosophy*, Oxford: Oxford University Press.

Hanahan, D. and Weinberg, R.A. 2011, "Hallmarks of Cancer: The Next Generation", *Cell*, 144, 646-74, doi: 10.1016/j.cell.2011.02.013.

Henderson, L. 2020, "The Problem of Induction", Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy (Spring 2020)*, https://plato.stanford.edu/archives/spr2020/entries/induction-problem.

Henschen, T. 2021, "How Strong is the Argument from Inductive Risk?", *European Journal for Philosophy of Science*, 11, 1-23, doi: 10.1007/s13194-021-00409-x.

Horton, R. 2020, "Offline: COVID-19 is Not a Pandemic", *The Lancet*, 396, 874, doi: 10.1016/S0140-6736(20)32000-6.

Islam, N., Lacey, B., Shabnam, S., Erzurumluoglu, A.M., Dambha-Miller, H., Chowell, G. et al. 2021, "Social Inequality and the Syndemic of Chronic Disease and COVID-19: County-level Analysis in the USA", *Journal of Epidemiology and Community Health*, 75, 496-500, doi: 10.1136/jech-2020-215626.

Koonin, E.V. and Dolja, V.V. 2013, "A Virocentric Perspective on the Evolution of Life", *Current Opinion in Virology*, 3, 546-57, doi: 10.1016/j.coviro.2013.06.008.

Laland, K.N., Sterelny, K., Odling-Smee, J., Hoppitt, W., and Uller, T. 2011, "Cause and Effect in Biology Revisited: Is Mayr's Proximate-Ultimate Dichotomy Still Useful?", *Science*, 334, 1512-16, doi: 10.1126/science.1210879.

Leach, M., MacGregor, H., Scoones, I., and Wilkinson, A. 2021, "Post-Pandemic Transformations: How and Why COVID-19 Requires Us to Rethink Development", *World Development*, 138, 105233, doi: 10.1016/j.worlddev.2020.105233.

Mayr, E.1961, "Cause and Effect in Biology", *Science*, 134, 1501.

McGwin, G. 2010, "Causation in Epidemiology", *American Journal of Ophthalmology*, 150, 599-601, doi: 10.1016/j.ajo.2010.06.031.

Morens, D.M. and Fauci, A.S. 2020, "Emerging Pandemic Diseases: How We Got to COVID-19", *Cell*, doi: 10.1016/j.cell.2020.08.021.

McMahon, N.E. 2021, "Understanding COVID-19 through the Lens of 'Syndemic Vulnerability': Possibilities and Challenges", *International Journal of Health Promotion and Education*, 59, 67-69, doi: 10.1080/14635240.2021.1893934.

Murányi, A. and Varga, B. 2021, "Relationship Between the COVID-19 Pandemic and Ecological, Economic, and Social Conditions", *Frontiers in Public Health*, 9, 1-10, doi: 10.3389/fpubh.2021.69419.

Randolph, D.G. et al. 2020, *Preventing the Next Pandemic: Zoonotic Diseases and How to Break the Chain of Transmission*, Nairobi, Kenya: United Nations Environment Programme and International Livestock Research Institute, https://www.unep.org/resources/report/preventing-future-zoonotic-disease-outbreaks-protecting-environment-animals-and (accessed November 11, 2021).

Paul, G. 2021, "Coronavirus Disease 2019 as a Syndemic", *Reviews in Medical Virology*, 31, 1-3, doi: 10.1002/rmv.2212

Pescaroli, G., Galbusera, L., Cardarilli, M., Giannopoulos, G., and Alexander, D. 2021, "Linking Healthcare and Societal Resilience During the Covid-19 Pandemic", *Safety Science*, 140, 105291, doi: 10.1016/j.ssci.2021.105291.

Potochnik, A. 2017, *Idealization and the Aims of Science*, Chicago: University of Chicago Press.

Raboisson, D. and Lhermie, G. 2020, "Living With COVID-19: A Systemic and Multi-Criteria Approach to Enact Evidence-Based Health Policy", *Frontiers in Public Health*, 8, 1-7, doi: 10.3389/fpubh.2020.00294.

Salmon, W.C. 1989, *Four Decades of Scientific Explanation*, Minneapolis: University of Minnesota Press.

Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D. et al. 2019, "How to Make Ecological Models Useful for Environmental Management", *Ecological Modelling*, 411, 108784, doi: 10.1016/j.ecolmodel.2019.108784.

Shimonovich, M., Pearce, A., Thomson, H., Keyes, K. and Katikireddi, S.V. 2020, "Assessing Causality in Epidemiology: Revisiting Bradford Hill to Incorporate Developments in Causal Thinking", *European Journal of Epidemiology*, 0123456789, doi: 10.1007/s10654-020-00703-7.

Shmueli, G. 2010, "To Explain or to Predict?", *Statistical Science*, 25, 289-310, doi: 10.1214/10-STS330.

Singer, B.J., Thompson, R.N. and Bonsall, M.B. 2021, "The Effect of the Definition of 'Pandemic' on Quantitative Assessments of Infectious Disease Outbreak Risk", *Scientific Reports*, 11, 1-13, doi: 10.1038/s41598-021-81814-3.

Singer, M., Bulled, N., Ostrach, B., and Mendenhall, E. 2017, "Syndemics and the Biosocial Conception of Health", *The Lancet*, 389, 941-50, doi: 10.1016/S0140-6736(17)30003-X.

Sironi, M., Hasnain, S.E., Rosenthal, B., Phan, T., and Luciani, F. 2020, "SARS-CoV-2 and COVID-19: A Genetic, Epidemiological, and Evolutionary Perspective", *Infection, Genetics and Evolution*, 84, 104384.

Taylor, L.H., Latham, S.M. and Woolhouse, M.E.J. 2001, "Risk Factors for Human Disease Emergence", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356, 983-89, doi: 10.1098/rstb.2001.0888.

Vineis, P. 2003, "Causality in Epidemiology", *Sozial- Und Praventivmedizin*, 48, 80-87, doi: 10.1007/s00038-003-1029-7.

Vineis, P. and Kriebel, D. 2006, "Causal Models in Epidemiology: Past Inheritance and Genetic Future", *Environmental Health: A Global Access Science Source*, 5, 1-10, doi: 10.1186/1476-069X-5-21.

Zabaniotou, A. 2020, "A Systemic Approach to Resilience and Ecological Sustainability During the COVID-19 Pandemic: Human, Societal, and Ecological Health as a System-Wide Emergent Property in the Anthropocene", *Global Transitions*, 2, 116-26, doi: 10.1016/j.glt.2020.06.002.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. 2020, "A Novel Coronavirus from Patients with Pneumonia in China, 2019", *New England Journal of Medicine*, 382, 727-33, doi: 10.1056/nejmoa2001017.

Vaccination Uptake Interventions: An EBM+ Approach

Daniel Auker-Howlett and Jon Williamson

Department of Philosophy and Centre for Reasoning University of Kent

Abstract

As the COVID-19 pandemic has demonstrated, barriers to vaccination uptake are heterogeneous and vary according to the local context. We argue that a more systematic consideration of local social and behavioural mechanisms could improve the development, assessment and refinement of vaccination uptake interventions. The EBM+ approach to evidence appraisal, which is a development of a recent line of work on the epistemology of causality, provides a means to evaluate mechanistic studies and their role in assessing the effectiveness of an intervention. We argue that an EBM+ methodology offers several potential benefits for research on vaccination uptake interventions. It also motivates the use of detailed mechanistic models, rather than the high-level logic models used by process evaluations, for example.

Keywords: Vaccination uptake interventions, Methodology, Evidence-based medicine, Mechanism, Mechanistic models; EBM+, Behavioural science.

1. Introduction

Immunisation is an integral part of global healthcare provision. It has helped to drive a massive reduction in worldwide annual child (under the age of 5) mortality, from 9.6 million in 2000 to 5.4 million in 2017 (WHO 2013; UNICEF 2018). It is estimated that annual deaths from just 5 vaccine-preventable diseases (diphtheria, measles, neonatal tetanus, pertussis and poliomyelitis) have dropped by 0.5 million a year since 2010. Vaccination coverage is one way of continuing to progress these achievements. There are licensed vaccines for 27 diseases, and to be licensed requires demonstration of efficacy. But effectiveness depends on much more than whether the vaccine elicits an appropriate immune response and protects the immunised from disease—vaccination coverage must also reach sufficient levels. The WHO's Global Vaccine Action Plan (GVAP) sets out a target of 90% coverage at the national level and 80% in every district by 2020. While coverage for most vaccines has substantially increased, many GVAP targets were not met. For example, global coverage for the 2nd dose of a measles vaccine has in-

ISSN 2465-2334

DOI 10.14275/2465-2334/202113.auk

creased by approximately 2/3rds, but absolute levels are still below 70% (MacDonald et al. 2020). If not enough people are being vaccinated, then immunisation programmes will fail.

Coverage depends on two broad sets of requirements. A sufficient stock of vaccines and the capacity to administer the vaccine to the whole target population are examples of *supply* requirements. Problems of supply result from limitations of infrastructure and resourcing. Accordingly, they are dealt with by approaches that focus on efforts to obtain sufficient resources and improve political will, e.g., investment in manufacturing and international aid. Even with sufficient supply, vaccination coverage may still fail to ensure population immunity, which can be explained by a number of factors that are relevant to the *demand* for vaccines. Problems of demand result from a wide variety of factors, including beliefs about vaccine safety, efficacy and utility. Interventions to increase demand for vaccines tend to focus on changing individual—and sometimes societal—beliefs and values.

One way to increase demand is to apply behavioural science. The Behavioural Insights Team, for example, argue that using psychological, sociological and related research to change vaccination behaviours is an avenue with much potential to increase demand for vaccination (Merriam and Behrendt, 2020). The World Health Organisation (WHO) have endorsed this strategy. Behavioural insights also play a major role in the response to COVID-19—see WHO 2020 and Betsch et al. 2020. Having obtained a sufficient supply of an efficacious vaccine, the focus shifts to interventions on behaviour to ensure there is sufficient uptake for the vaccine to be effective.

For the many infectious diseases that are the target of vaccination programmes worldwide, the methodology used to guide the development and assessment of interventions is crucial. In recent decades, the focus of efforts to increase vaccination coverage has been on low to middle income countries (LMICs) on whom the burden of infectious disease falls most heavily. COVID-19 has exposed how the consequences of getting vaccination programmes right can affect countries across the economic spectrum. In LMICs it is particularly important not to devote limited resources to vaccination programmes unless they are likely to have enough uptake to be effective. On the other hand, while high income countries (HICs) may be able to devote resources to vaccination programmes, lack of uptake threatens to hinder any progress made against an infectious disease that for the first time in half a century poses a real and present danger to the health and economy of HICs. This paper argues that the development and assessment of interventions to increase vaccination uptake would benefit from changes to methodology motivated by the EBM+ programme. As explained in §3, EBM+ emphasises the importance of mechanistic evidence when assessing causal claims. Here, the causal claims of interest are claims about the effectiveness of vaccine uptake interventions.

EBM+ is a development of the recent mechanistic turn in the philosophy of science. Russo and Williamson 2007 argued that in order to establish a causal claim in medicine one needs to establish that the putative cause and effect are correlated and that they are linked by some mechanism that can account for this correlation. If correct, this suggests that present-day evidence-based medicine (EBM), which focusses on clinical studies to the exclusion of mechanistic studies, may be overlooking important evidence (Williamson 2019). EBM+ augments EBM with methods for properly assessing mechanistic studies and integrating these assessments with those of clinical studies in order to assess causation (Parkkinen et al. 2018).[1]

EBM+ therefore also has important consequences for the use of models in establishing causal claims. In particular, well-confirmed mechanistic models can help to establish the existence of a linking mechanism, thereby confirming a causal claim of interest. Thus mechanistic models can be useful when establishing the effectiveness of vaccine uptake interventions. This suggests a greater role for mechanistic models than, say, the logic models of process evaluations, which are currently used to assess vaccine uptake interventions.

In §2 we describe the current methodology for assessing effectiveness and argue that it has certain limitations. We present the alternative EBM+ approach in §3. We then develop two case studies of the use of EBM+. In §4 we consider an example in which EBM+ would deem evidence of effectiveness to be weak and in §5 an example in which evidence of effectiveness is strong. We argue that each case would benefit from an EBM+ approach. We conclude in §6 that an EBM+ approach has much to offer vaccination uptake research.

2. The Status Quo

The dominant methodology for the assessment of interventions to increase vaccination uptake is that of the standard approach to assessment in the health sciences, namely evidence-based medicine (EBM). This methodology prioritises evidence obtained by association studies—particularly randomised controlled trials (RCTs)—when assessing the effectiveness of interventions, and downplays the evidential role of mechanistic studies. An association study of a vaccination uptake intervention tests whether the intervention is associated with uptake, and usually also ascertains the extent of any observed correlation between the two. On the other hand, a mechanistic study aims to shed light on features of the complex of mechanisms by which the intervention might influence uptake, including the variables that are intermediate between cause and effect and the entities and activities involved in the mechanisms and their organisation. According to EBM, mechanistic evidence may help to suggest a new intervention, but it provides at best very weak evidence of effectiveness. Thus mechanistic evidence is rarely considered by EBM-based systematic reviews of effectiveness (Williamson 2019: §1.3).

Vaccination uptake interventions, in particular, follow present-day EBM, which deems mechanistic evidence relevant to the context of discovery (i.e., hypothesising the intervention) but not the context of justification (i.e., assessing effectiveness). For example, behavioural science is used to suggest interventions. These interventions may exploit particular cognitive biases that have been identified by theoretical psychology. For instance, omission bias is the tendency for people to judge harmful actions more harshly than inaction, even where both cause equivalent harm (Merriam and Behrendt 2020: 13). There is some evidence that omission bias plays a part in a mechanism that influences vaccination attitudes in the US. An intervention may thus be proposed to target omission bias. This process is analogous to the way in which pharmaceutical interventions are suggested by 'basic science' research. Methods of the biomedical sciences are used

[1] EBM+ is not without its critics. See Williamson 2019: §1 and references therein for further discussion.

to identify features of mechanisms of action of potential pharmaceuticals. Those that show promise are then tested in clinical trials.

That the current assessment of the effectiveness of vaccination uptake interventions favours association studies over mechanistic studies is witnessed by the fact that systematic reviews of vaccination uptake interventions typically only include evidence obtained in RCTs (Manakongtreecheep 2017; Jacobson Vann et al. 2018; Merriam and Behrendt 2020). Moreover, standard EBM evaluative frameworks are adapted and used to evaluate the quality of the evidence for these interventions. For example, Merriam and Behrendt 2020 use the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) framework to evaluate the quality of the studies included in their review. GRADE focusses on association studies.

It is clear that current methodology downplays or outright excludes mechanistic evidence from the assessment of vaccination uptake interventions. There are however some exceptions. First, the WHO 'tailoring vaccination programmes' guidance instructs designers of programmes to refine interventions to take account of barriers to, and facilitators of, vaccination (WHO Europe 2013). Second, a recent move to emphasise 'theory' in the design and evaluation of behavioural change interventions, where theory is defined as a "set of analytical principles or statements designed to structure our observation, understanding and explanation of the world" (Moore et al. 2019: 3). This brings the importance of mechanisms to light, but is unlikely to account for all the mechanisms at work for an intervention in a specific context (Moore and Evans, 2017). Third, *process evaluations* seek to elucidate the causal assumptions of an intervention, and attempt to identify how an intervention works (Craig et al. 2019). Alongside the testing of factors important to the implementation of association studies, process evaluations look for the mechanisms relevant to an intervention's effectiveness. We will revisit process evaluation in §5.

While these efforts cannot be discounted, the failure of the wider field to systematically consider mechanistic evidence is a problem for several reasons. Firstly, extrapolating the results of research from one population to another benefits from a careful scrutiny of mechanisms. In particular the social and behavioural mechanisms that the mechanism of action of the intervention interacts with will differ between contexts. For example, educational barriers to vaccination in HICs primarily concern beliefs about safety or importance, whereas in LMICs access to information about the benefits of vaccines is the main educational barrier (Gardner et al. 2010; Sadaf et al. 2013). Developing an intervention in a LMIC that addresses safety without improving access to information about benefits would be misguided (Aronson et al. 2021: §5). Second, association studies can play only a limited role in identifying why an intervention succeeds or fails. Articulating and evaluating the mechanisms that impinge on whether an intervention brings about an effect can help here. For example, one intervention to increase Human Papillomavirus (HPV) vaccination coverage in teenage girls involves administering vaccines in schools. In an evaluation of such a programme in the USA, parental approval was required for vaccination, so their beliefs about the importance of vaccination may have accounted for low participation rates (Stubbs et al., 2014). Taking account of the features of this parent-mediated mechanism is one way to improve the effectiveness of school-based HPV vaccination programmes (Stubbs et al. 2014). Last but not least, association studies alone are typically not enough to warrant a causal conclusion. Only several concordant studies with the best designs and implementation can do this. When an evidence base fails to meet this standard—and it often does—high-quality evidence of mechanisms helps to establish a causal connection, as we discuss in the next section. Indeed, RCTs, often cited as the gold-standard kind of association study, may be viewed as undesirable on epistemic grounds (Worrall 2007) or because they are often costly and ethically questionable, and can lead to research biases (Ravallion 2020). Thus the use of mechanistic evidence promises to improve the development and assessment of vaccination uptake interventions. A new methodology for causal evaluation, EBM+, systematises the evaluation of mechanistic evidence. Next, we introduce this methodology, before moving on to showing how it can improve upon the methods employed in two very different examples of public health interventions.

3. EBM+

EBM+ adds to standard EBM the explicit evaluation of evidence obtained by mechanistic studies (Parkkinen et al. 2018).

Figure 1 portrays the role of evidence in establishing a causal claim, according to EBM+. Association studies can be used to directly test whether the putative cause and effect are correlated (evidential channel C_1 in Figure 1). However, correlation is insufficient for causation: while some observed correlations are causal, others are attributable to various kinds of study bias, to confounding, to various kinds of non-causal relationship, or even to mere coincidence. What is distinctive about a correlation that is attributable to causation is that there is some mechanism complex by which instances of the cause are responsible for instances of the effect and which can account for the extent of the observed correlation. Some association studies—notably large, well-conducted RCTs—can provide indirect evidence of the existence of such a mechanism (channel C_2), by reducing the probability that any observed correlation is due to confounding. But there is a more direct way to ascertain whether there is an underlying mechanism that can account for the correlation: posit key features of the mechanism and assess whether mechanistic studies show those features to be present (channels M_1 and M_2). If these mechanism features are well confirmed then this in turn can make the causal claim more plausible. (In certain circumstances, it can even make it more plausible that there is a genuine correlation—channel M_3.) This account of the epistemology of causation is sometimes called 'Evidential Pluralism', to distinguish it from a monistic account that focuses exclusively on association studies, as is the case with standard EBM.

Mechanistic studies investigate features of the complex of mechanisms linking cause to effect. This mechanism complex includes the mechanism of action, by which the cause directly contributes to the production (or prevention) of the effect, together with any mechanisms that counteract or enhance the influence on the effect attributable to the mechanism of action: together, these mechanisms can explain the net correlation between cause and effect. Where the causal claim is a claim about the effectiveness of a medical intervention, this mechanism complex might involve mechanisms responsible for the functioning of systems in the human body; the progression of disease; the metabolism of pharmaceuticals; the functioning of medical devices; the distribution of and access to the intervention; and compliance with the intervention. Thus, the relevant mechanisms can be biological, physiological, chemical, physical, infrastructural, social, behavioural and psychological.[2] Accordingly, the methodologies employed by mechanistic studies are very heterogeneous. The key point such studies have in common is that they provide evidence for specific mechanism hypotheses, which hypothesise key features of the mechanism complex linking the intervention to the outcome.

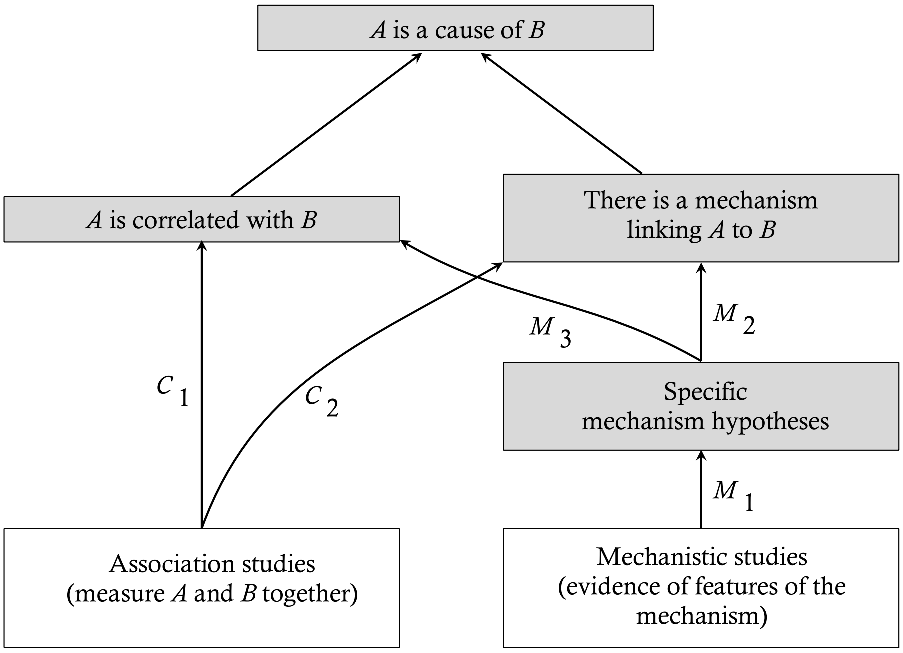

A is a cause of *B*
A is correlated with *B*
There is a mechanism linking *A* to *B*
M 3
M 2
C 1
C 2
Specific mechanism hypotheses
M 1
Association studies (measure *A* and *B* together)
Mechanistic studies (evidence of features of the mechanism)

Figure 1: Evidential relationships for assessing a causal claim (Williamson 2021).

In the present context, the outcome of interest is increased uptake of vaccination in a target population, so infrastructural, social, behavioural and psychological mechanisms are particularly important. For example, vaccination appointment reminders offer the potential to increase vaccination uptake (see §4). The mechanism of action is straightforward: a reminder prompts individuals to get their children or themselves vaccinated, which they then remember to do. This intervention intervenes on the 'pathology' of forgetting to have a vaccination, or of forgetting the correct vaccination date and time, and is supported by psychological mechanisms. Prospective memory is a type of memory that involves remembering to perform a specific action at a future time: in this case, remembering the date and time of vaccination. Many factors can decrease the reliability of prospective memory, including the emotional and motivational state of an individual (Jeong and Cranney 2009; Rendell et al. 2011; Schnitzspahn et al. 2014). This evidence supports the existence of the mechanism of action, whereby reminders intervene on an individual's memory of the vaccination appointment, which might have been diminished by external stressors affecting their emotional state and/or their motivation to procure vaccination. However, the mechanism of action interacts with a number of other mechanisms. For example, parents might be aware that the risk of death or serious illness from COVID-19 in children is

[2] See Kelly et al. 2014; Kelly and Russo 2018 for discussions of how many of these mechanisms can interact.

minimal (Viner et al. 2020). This information may make them believe that it is not worth their while to get their children vaccinated against COVID-19. This belief figures in a separate psychological mechanism operating concurrently with the mechanism of action and counteracting its effect. Evidence that this counteracting mechanism also operates in the target population will undermine confidence that the mechanism complex will lead to overall effectiveness. This is why it is important to consider various potential pathways in the mechanism complex, rather than solely the mechanism of action of the intervention.

Crucially, it is not enough to simply have a *story* of a mechanism—for EBM+, decisions must be based on evidence. According to EBM+, one needs to systematically evaluate the mechanistic studies relevant to those key features of the mechanism complex that are not already established by prior evidence. Thus, mechanistic studies are treated in the same way that association studies are treated by standard EBM. EBM+ provides methods for the systematic review of mechanistic studies, and guidance for combining this evidence with evidence from association studies to make a judgement about the plausibility of causality. See Parkkinen et al. 2018 for a detailed guide to these methods, and Auker-Howlett 2020: Ch.3 for an example of EBM+ applied to an evaluation of a pharmaceutical intervention on Middle East respiratory syndrome (MERS). Here, we shall just note some general features of the EBM+ evaluation process.

The first task for an EBM+ evaluation is to assess the association studies to determine whether they establish the existence of an appropriate correlation and an appropriate mechanism (evidential channels C_1 and C_2 in Figure 1). See Williamson 2019 for a discussion of what counts as 'appropriate' here. Existing evaluation techniques (e.g., 'GRADE') can be applied. If correlation and mechanism are both established then causation is established.

If a correlation has been found but it is not clear that this correlation is causal—i.e., attributable to an underlying mechanism—then the next step is assessing the quality of the individual mechanistic studies. To be rated as high quality, the methods used in the studies must be well understood, the experimental system must be similar to the target system, and the methods must be implemented properly.

Then one needs to consider whether the mechanistic studies establish the key features of the mechanism complex (M_1).[3] This is done on different grounds: do we have multiple studies showing consistent results across similar and different kinds of methods? In effect, we are checking here for a kind of robustness of results to changes in background conditions.

If evidence is missing for key features of the posited mechanism, or the mechanism could not plausibly account for the size of the observed correlation, then one's confidence in the mechanism is undermined (M_2). On the other hand, one's confidence may be raised if all the key features of a mechanism are confirmed in sufficient detail and if the mechanism can account for the correlation and its size. Only those mechanistic evidence bases that are high quality and support high confidence in a mechanism claim can establish the existence of a suitable mechanism complex. The status of a mechanism complex is then a function of both the quality of evidence and one's confidence in the mechanism complex. For example, the status is *established* when high quality evidence warrants high confidence

[3] See Parkkinen et al. 2018: 83-84, and Steel 2008 on what constitute 'key features' of a mechanism.

in the claim but only *provisionally established* when moderate quality evidence warrants high confidence. Lower status levels include *arguable*, *speculative*, *arguably false*, *provisionally ruled out* and *ruled out* (Parkkinen et al. 2018: 27).

The challenge for research on vaccination uptake is to articulate, evaluate and integrate mechanistic evidence, in order to design and assess interventions. EBM+ offers a method for systematising this process. In the next two sections we analyse two case studies to show how the EBM+ methodology can be employed to improve current practice.

4. SMS Reminders and Cash Incentives for Increased Vaccination Coverage

Our first case study is one in which the evidence base has significant deficiencies.

4.1 The Evidence Base and its Limitations

As noted above, one kind of intervention to boost vaccination demand involves *reminders*: these take the form of phone, text or email reminders sent to the individuals to be vaccinated or to parents of children to be vaccinated. Vaccine reminders have been extensively investigated as an intervention in HICs. A Cochrane review including a meta-analysis of 55 studies found that reminders increased vaccination rates by 8% (Jacobson Vann et al. 2018). What evidence there is arising from association studies in LMICs has been reviewed by Merriam and Behrendt 2020, who conclude that reminders are 'generally effective'. However, both evidence bases are beset by numerous problems with the quality of evidence: there are a number of defects of the evidence base.

In the LMIC evidence base, variability of intervention is a problem. Indeed, some studies implement a combination of interventions. For example, a large RCT in Kenya supposedly demonstrates the effectiveness of SMS reminders, but in fact it is SMS reminders plus cash incentives that are associated with increased vaccination uptake (Gibson et al. 2017). A trial arm testing only SMS reminders displayed no increase in uptake relative to controls. There is another form of variability of intervention: different kinds of reminders. For example, another RCT, this time in Nigeria, found that phone call reminders were effective compared with a training programme for health care workers (Brown et al. 2016). This result was corroborated by Ekhaguere et al. 2019. Thus the evidence can be interpreted as supporting the effectiveness of phone call reminders, but not SMS reminders. The small size of the LMIC evidence base exacerbates this problem, while the larger HIC evidence base for each kind of reminder may ameliorate the issue.

A second defect of the LMIC evidence base concerns specificity of outcome. The conclusion of Gibson et al. 2017 was that SMS messages together with cash incentives were effective at increasing vaccination coverage. However, the results of the study do not support this general conclusion. Vaccination coverage here refers to 'full immunisation' of 8 vaccines, yet only for the measles vaccine was there a significant increase in coverage. Baseline coverage—measured for the control group—was near 100% for the other 7 vaccines. An increase from 87% to 90% in the measles group boosted the average. This pattern of results is replicated by Ekhaguere et al. 2019. A more precise claim is that reminders are effective for increasing measles vaccine uptake. This is no mean feat of course, but in these cases the claim should have been made specific to the vaccine, rather than to vaccination in general. Compounding this problem is the high baseline coverage in the studies of Gibson et al. 2017 and Ekhaguere et al. 2019. Interventions on vaccination uptake are motivated by the fact that baseline coverage is generally below levels sufficient to ensure population immunity. Recall from §1 that the WHO GVAP sets a target of 90% coverage for all vaccines (WHO 2013). In the study of Gibson et al. 2017, even the measles vaccine baseline was 87%, and the remaining 7 vaccines ranged from 96% to 98%. The authors hypothesised that this was due to the high rate of study dropouts, consisting of the poorest, most mobile and youngest mothers. Plausibly, such people are those who are the primary targets of uptake interventions. Thus the sample tested in this study appears to be highly unrepresentative of the target population. The results of Ekhaguere et al. 2019 are less worrisome. Although only coverage for measles and the third of three doses of pentavalent vaccine (a combination vaccine for diphtheria, tetanus, whooping cough, polio, and Haemophilus influenzae type b disease) is lower than the 90% target, the dropout rate was only 8%. So in the context of those specific vaccinations the reminders may still be effective. Thus, a third defect of the evidence base concerns the representativeness of the sample.

A fourth defect that affects both evidence bases arises from a lack of blinding of participants or researchers to trial arm allocation. For example, Haji et al. 2016 reported that researchers checked up on one trial arm whose participants were given stickers to remind them of vaccination dates. This would have involved researchers knowing which trial arm the participant was in and opens the possibility that these interactions influenced outcomes—a form of bias. Performance bias and detection bias both result from lack of blinding of trial personnel to study group allocations. Jacobson Vann et al. 2018 rates 28% of studies conducted in HICs to be at low risk of performance bias, 66.7% at unclear risk, and 5% at high risk. For detection bias, figures are 29.3% low, 68% unclear, 2.7% high. One example of a study at low risk of both kinds of bias is an RCT on the effect of reminder/recalls (Szilagyi et al. 2013). While trial personnel were indeed blinded to study group allocations, trial participants were not. The review does not consider the kinds of biases resultant from lack of blinding of trial participants. The problem is that it is plausible that the participants' knowledge of being in an experiment may be what influences vaccination uptake, rather than the intervention itself. Worse still, there may be a general inability to blind participants in studies in the social sciences (Cartwright and Deaton 2016). So this defect may be unavoidable for vaccination uptake research.

4.2 An EBM+ Perspective

A closer scrutiny of the LMIC evidence base reveals that the evidence base only slightly supports the effectiveness of reminders, and that this is only for phone call reminders and only for the measles vaccine. A review of specific kinds of reminders, or studies that test single vaccines, might be more informative. This recommendation accounts for defects 1-3 and is in line with the EBM perspective: the standard response to defects in the evidence base is to demand more and better trials to be carried out. Large scale trials are costly and time consuming, but funders may see the worth of carrying out trials if vaccination coverage can be boosted. However, even if resources can be committed to more trials, it is not clear that the EBM perspective can resolve the issues presented here. The fourth defect is something quite general that might limit all trials performed in this context—it is very difficult to come up with "placebo" reminders.

From the perspective of EBM+, it seems plausible that the studies establish that a correlation exists between reminders and vaccine uptake. The errors that thwart a causal conclusion could be explained away if one could establish that an appropriate mechanism complex exists linking reminders and vaccination uptake. The basic logic of an EBM+ causal evaluation would then lead us to attribute the correlation to a causal relationship. Let us first apply this logic to the reminder case, and then advance some reasons why this improves on the EBM perspective. Take for example Haji et al. 2016, where researchers might have known who was in each group. One plausible explanation of the observed correlation is that researchers knowingly or unknowingly influenced participants in the SMS reminder group to get vaccinated. Were we however to establish that there is a mechanism of action linking SMS reminders to vaccination uptake, and that other mechanisms in the target population could not fully counteract the mechanism of action (see §3), then this would warrant greater confidence in the effectiveness of the intervention in the target population.

Note that, on both the EBM and EBM+ perspectives, low-quality clinical trials fail to establish *causation*, but on the EBM+ perspective such trials may still provide high-quality evidence of *correlation*. Similarly, mechanistic evidence does not normally suffice to establish causation on either perspective, but EBM+ treats this kind of evidence seriously, and can evaluate it as high-quality evidence of *mechanism*. Taken together, such evidence can be high-quality evidence for causation. Thus, one reason why the EBM+ perspective improves on standard EBM is that it appeals to different kinds of studies, which can reinforce one another. The standard problems that make it hard to infer causation just from mechanistic studies (the complexity of mechanisms, and the presence of counteracting mechanisms, both of which makes it hard to establish a net correlation) are different to those that make it hard to infer causation just from association studies (bias, confounding, statistical blips, non-causal connections). Establishing a correlation is just what is required to overcome the limitations of mechanistic studies, and establishing the existence of a mechanism is just what is required to overcome the limitations of association studies. Thus, by considering both sorts of study together, EBM+ can avoid the pitfalls of each.

The EBM+ approach can also enable quicker decisions, which may be vital in resource- and time-limited contexts such as vaccination uptake research. Repeating trials takes time and money. We have recently seen the need for timeliness in COVID19 vaccination research, and, as Aronson et al. 2021 show, a consideration of the mechanisms at play can lead to a more conclusive evidence base, obviating the need for further association studies. Mechanistic studies can be carried out concurrently with, or very often before, association studies. A causal evaluation including both trial and mechanistic evidence can then be carried out at the close of the trials.

It is important to note that the problem with the current research on reminders is a lack of evidence of mechanisms—not evidence that there is no mechanism. This means that the effectiveness claim is still open, and that more research into relevant mechanisms may well help. An EBM+ approach therefore helps to identify the gaps in the evidence base which need to be filled by commissioning further research. Another advance that EBM+ offers in this area is in the *evaluation* of mechanistic evidence. It is not enough to just *conduct* mechanistic studies. A systematic approach to evaluating evidence is also needed—this is something that both EBM and EBM+ agree on.

In the next section we identify some methods currently used in public health interventions to evaluate mechanisms, and how EBM+ may improve on these methods.

5. The Welsh National Exercise Referral Scheme

Our second case study is an example of a public health intervention for which there is stronger mechanistic evidence. Although this case is not related to vaccination, it is instructive because it helps to show what constitutes a strong evidence base, as well as the usefulness of EBM+ in framing the assessment of the evidence base.

5.1 The Evidence Base

Increased physical activity is an important means to reduce chronic disease. The Welsh national exercise referral scheme (NERS) implements an intervention to increase physical activity, namely *exercise referral*. Exercise referral schemes (ERS) "typically [...] involve health professional referral to a leisure facility, agreement of an exercise programme with an instructor, and discounted access to leisure facilities for 10–12 weeks" (Littlecott et al. 2014: 2). The study that assessed NERS was a pragmatic RCT (Murphy et al. 2012) which found that "the intervention was associated with significant improvements in physical activity for patients referred with coronary heart disease risk factors (though not for patients referred for mental health reasons)" (Littlecott et al. 2014: 2).

However, ERS have in general seen little long-term impact on physical activity. This may be explained by a number of factors: ERS are heterogeneous in design; they differ in their mechanisms of action (e.g., one intervention may increase social support by forming new social groups in class-based exercise, while another directly targets an individual's motivation); and demographic factors and health conditions can each affect physical activity. Evaluating the mechanisms at work in particular schemes is a way to explain whether, why and how the intervention worked.

It is now recommended that process evaluations be used to evaluate complex interventions (Craig et al. 2008; Moore and Evans 2017; Craig et al. 2019). The primary goal of this kind of evaluation is to refine the implementation of the intervention, by taking account of the social and behavioural mechanisms that the mechanism of action interacts with. For NERS, to understand the theoretical assumptions being made by the design of the intervention, and the mechanisms by which the intervention brings about the effect, a mixed methods process evaluation was undertaken (Moore et al. 2013; Littlecott et al. 2014). This evaluation used quantitative and qualitative studies to identify key psychosocial mediators as well as factors that influenced whether the intervention was effective or not.

A qualitative study used interviews to "explor[e] patients' motivations for attending NERS, their opinions of the scheme, perceived impacts, mechanisms of change, barriers and facilitators of attendance and future exercise intentions" (Moore et al. 2013: 483). This study found that important factors for success were: effective professional supervision and guidance; having the social support of other patients; and the range of classes, times and locations. Such factors are key components of mechanisms crucial to how NERS brought about its effect. For instance, having adequate social support was found to be crucial to adherence in NERS. But social support was often dependent on having a referral from a physician. This was because the participant's family were more willing to support their efforts when referral came from a health authority, as the scheme was perceived as important for improving more than just general fitness. So the social support mechanism also includes referral from a physician. Psychological theory also suggests a number of mechanisms by which NERS brings about its effects. Change in activity is hypothesised to occur when individuals have high levels of autonomous motivation, e.g., when they find an activity enjoyable. Indeed, autonomous motivation is associated with increased physical activity. When individuals see some behaviour change as an effective means of achieving desired outcomes, and as within their capabilities, they are more likely to enact that change. This 'self-efficacy' mechanism has again been associated with increased levels of physical activity.

A quantitative study tested whether referral to NERS was associated with effects on these mechanisms at 6 months, and whether impacts on physical activity were mediated by change in the mechanisms at 12 months. To test for effects on these hypothesised mediators at 6 months, participants were interviewed and effects were assessed by regression tests. Assessing whether a variable mediates change in physical activity involved statistical tests that looked for: (i) whether the proposed mediator is associated with the outcome (by separately calculating estimates for the effects of the intervention and mediator on physical activity while adjusting for the other variable), and (ii) what proportion of the total effect is explained by indirect effects (Littlecott et al. 2014: 4-5). This assessment found significant effects for autonomous motivation and social support for exercise, but none for self-efficacy. The authors conclude that the intervention's effect on exercise activity is mediated by autonomous motivation, and that the findings of this analysis support the use of self-determination theory as a framework for development and implementation of the exercise referral scheme.

5.2 An EBM+ Perspective

In some ways, process evaluations align with EBM+ methodology. 'Theory' is being used to suggest specific mechanism hypotheses: psychosocial mediators are features of the mechanisms. The studies then provide evidence for the existence of these features, and for the existence of a mechanism complex that accounts for the correlation—in EBM+ parlance they are mechanistic studies. Additionally, the combination of the results of the association studies and the process evaluation is viewed as more informative than the results of the association studies alone. Thus, the use of the process evaluation demonstrates that it is feasible to consider mechanistic evidence when assessing complex public health interventions. However, a full EBM+ approach improves upon current methodology on two fronts.

Firstly, a process evaluation goes as far as testing features of mechanism hypotheses in mechanistic studies, but EBM+ goes further in requiring a systematic evaluation of evidence generated by these studies. Systematic reviews of the evidence obtained in association studies are commonplace in vaccination and public health research. EBM+ does not differ from EBM in this respect and provides its own methods for the evaluation of mechanistic evidence (Parkkinen et al. 2018). This improves upon current methodology, because it is not enough merely to have some mechanistic evidence—that evidence must also be high quality. As described in §3, Parkkinen et al. 2018 provide guidance for evaluating the quality of mechanistic evidence and for integrating such evaluations with association study evidence to assess causal claims.

On its own, a process evaluation is less able to establish causality. The main difficulty is in establishing the existence of an appropriate mechanism. This is unsurprising, as the core goal of a process evaluation is to identify whether the intervention has been implemented correctly. The elucidation of relevant mechanisms is important for achieving this task, but the central goal of a process evaluation is not to articulate the mechanism of action in order to confirm causation. However, the studies conducted in the process evaluation for NERS would come out as providing strong mechanistic evidence according to an EBM+ evaluation: the experimental and target systems are almost identical; the methods are well established in public health research; and, both qualitative and quantitative methods identified autonomous motivation as a feature of the mechanism complex, thus demonstrating robustness of results. While the evidence is strong, it is not clear that any mechanism is definitively established—there are too many un-evidenced gaps in the mechanism complex. However, studies conducted outside of the process evaluation could also provide evidence for the relevant mechanisms at work in NERS. For example, Littlecott et al. (2014: 7) note that "improvements in autonomous motivation after attendance at an exercise referral scheme have been described by a number of previous studies" and provide one example: Markland and Tobin 2010. But only a full systematic evaluation of all the relevant evidence would allow those extraneous studies to bear on the status of the relevant mechanisms. This is something that EBM+ but not a process evaluation offers. The detail provided by an EBM+ evaluation evidently goes beyond the detail provided by the process evaluation.

Additionally, EBM+ benefits from its appeal to the *concept* of mechanism. Process evaluations talk of 'theory' and 'mediators', but it is not made clear that what is elucidated is a mechanism complex. Talk of 'facilitators' and 'barriers' in vaccination research is also rather impoverished, although perhaps more suggestive. On the other hand, EBM+ draws on a rich seam of work in the philosophy of science on how best to characterise mechanisms, e.g., Machamer et al. 2000; Illari and Williamson 2012; Craver and Darden 2013. Mechanisms decompose into component entities, activities and organisational features and are connected by processes. This richer characterisation helps because it is clearer where each feature acts in a causal pathway. In the NERS example, individuals have varying degrees of *autonomous motivation*, which is credited with playing a mediating role on the effectiveness of the programme. This psychosocial mediator is thus a feature of a mechanism, but it is clear that it is a component of a mechanism consisting of a much richer structure. Thinking about how this component acts on other components, and how that sequence is organised, facilitates a description of a testable mechanism hypothesis. One way in which this method is better than thinking in terms of 'mediators' of a causal pathway is that it includes organisational information.

Organisation is key to understanding the mechanism, and it is baked into the characterisation of mechanisms used by EBM+. Another benefit of a richer characterisation is that it helps to identify other features that are crucial to whether the mechanism operates. For example, identification of two linked entities in a mechanism and no activity between them implies that an activity is missing from the mechanism description. Thus, the appeal to the concept of mechanism helps to identify gaps in the evidence base. This kind of reasoning is used widely in mechanism discovery (see, e.g., Darden and Craver 2002; Craver and Darden 2013), but it is also useful when evaluating the plausibility of mechanisms in intervention assessment—an evaluation of an incorrect characterisation of a mechanism hypothesis is evidently impoverished.

6. Conclusion

Aronson et al. 2021 suggested that an EBM+ approach might benefit several areas of COVID-19 research, including COVID-19 vaccination research. In this paper we have developed the case for an EBM+ approach to vaccination uptake research. The COVID-19 pandemic has demonstrated the problems that a mixture of vaccine hesitancy and logistical difficulties pose for vaccine roll-out. The EBM+ approach to intervention assessment is readily applicable to vaccine uptake interventions, because the barriers and facilitators to vaccination depend on local social and behavioural mechanisms. A mechanistic perspective leads to systematic scrutiny of mechanistic studies—and mechanistic models—in addition to association studies, and this broader evidence base can lead to better judgements of effectiveness. The mechanistic perspective can also aid the development and refinement of vaccination uptake interventions, and their extrapolation to new populations.

The foregoing analysis presents a positive picture of EBM+ compared with current public health intervention research: it provides a systematically evaluated evidence base supporting more accurate effectiveness judgements, and all grounded in a richer conceptual framework. The benefits of more accurate effectiveness judgements were discussed in §4, but the fuller picture presented here has further implications for vaccination research. Consider the 'increasing vaccination model' in Figure 2. The authors of the model constructed it to represent the multiple strategies for increasing vaccination uptake, and Merriam and Behrendt 2020 use it as a basis for conceptually organising different kinds of vaccination uptake research. However, this model would benefit from the kind of EBM+ treatment the analysis in §5 applied to process evaluations. In one respect, the model is a representation of a very high-level mechanism, e.g., "what people think and feel" is a feature of a mechanism hypothesis that affects a "motivation" feature. However, this mechanism will in reality be much richer, and the mechanistic models employed by EBM+ can capture this. Moving to an EBM+ approach would thus improve the 'increasing vaccination model' by increasing its accuracy. For example, 'reminder' interventions are categorised by Merriam and Behrendt 2020 as a set of interventions that deal with passive under-vaccination, which arises from ambivalence, uncertainty, or logistical issues. Factors that contribute to passive under-vaccination fall into the 'what people think and feel' element of the model. So reminders intervene on thoughts and feelings and bring about effects on 'motivation'. Brewer et al. 2017 suggest that successful vaccination uptake will often require employing multiple strategies, and the model makes sense of how they all fit together. One can categorise various interventions according to the model and work out which strategies they will be similar to, or which they will be causally downstream from. Yet without knowing the key details of the various mechanisms that reminders are intervening on, this model tells us little about how they may integrate with other strategies, e.g., cash transfers. In fact, without more detail, and as independent mechanisms often counteract one another, it is plausible that the mechanisms of action of the various vaccination uptake interventions counteract one another as well. If such models are to inform reasoning about vaccination uptake interventions, the detailed mechanistic models employed by EBM+ are a better option.

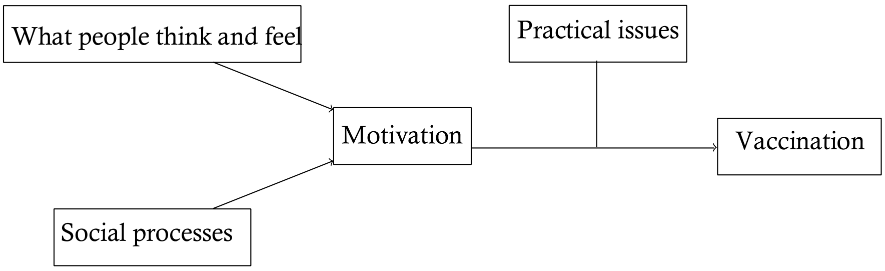


Figure 2: Increasing vaccination model, derived from Brewer et al. 2017.

With all that said, one might think that establishing mechanisms in complex public health interventions—such as COVID-19 vaccination uptake interventions—will simply be too difficult. Research into mechanisms in the biomedical sciences can isolate and manipulate systems in laboratory studies. Research on the behaviour of humans is more limited in this respect. So asking for mechanisms in order to establish effectiveness might seem to be setting the bar too high. However, as we noted above, behavioural mechanisms can be and have been established. Moreover, even where causation cannot be conclusively established, EBM+ nevertheless facilitates judgements about the relative plausibility of the effectiveness of several interventions. Relative plausibility is important, as recommendations on the widespread implementation of interventions need not be restricted to interventions for which effectiveness is established. Indeed, the GRADE system separates judgements of effectiveness from judgements of the adequacy for recommendation. Judging adequacy involves assessing benefits and costs of both intervening and failing to intervene (Andrews et al. 2013). If the evidence for the effectiveness of an intervention is less than conclusive, but there are few costs, then a recommendation for widespread implementation may yet be reasonable.[4]

References

Andrews, J.C., Schünemann, H.J., Oxman, A.D., Pottie, K., Meerpohl, J.J., Coello, P.A., Rind, D., Montori, V.M., Brito, J.P., Norris, S., Elbarbary, M., Post, P., Nasser, M., Shukla, V., Jaeschke, R., Brozek, J., Djulbegovic, B., and Guyatt, G. 2013, "GRADE Guidelines: 15. Going from Evidence to Recommendation—Determinants of a Recommendation's Direction and Strength", *Journal of Clinical Epidemiology*, 66, 7, 726-35.

[4] We are very grateful to the three anonymous referees for many helpful comments. DAH was funded by Arts and Humanities Research Council AH/L503861/1 and UK Research and Innovation Global Challenges Research Fund. JW was funded by Arts and Humanities Research Council AH/M005917/1 and Leverhulme Trust RPG-2019-059.

Aronson, J.K., Auker-Howlett, D., Ghiara, V., Kelly, M.P., and Williamson, J. 2021, "The Use of Mechanistic Reasoning in Assessing Coronavirus Interventions", *Journal of Evaluation in Clinical Practice*, 27, 684-93.

Auker-Howlett, D. 2020, *Evidence Evaluation and the Epistemology of Causality in Medicine*, Doctoral dissertation, University of Kent.

Betsch, C., Wieler, L.H., and Habersaat, K. 2020, "Monitoring Behavioural Insights Related to COVID-19", *The Lancet*, 395, 10232, 1255-56.

Brewer, N.T., Chapman, G.B., Rothman, A.J., Leask, J., and Kempe, A. 2017, "Increasing Vaccination: Putting Psychological Science into Action", *Psychological Science in the Public Interest*, 18, 3, 149-207.

Brown, V.B., Oluwatosin, O.A., Akinyemi, J.O., and Adeyemo, A.A. 2016, "Effects of Community Health Nurse-Led Intervention on Childhood Routine Immunization Completion in Primary Health Care Centers in Ibadan, Nigeria", *Journal of Community Health*, 41, 2, 265-73.

Cartwright, N. and Deaton, A. 2016, "Understanding and Misunderstanding Randomized Controlled Trials", CHESS Working Paper No. 2016-05, 44, 667526, 1-63.

Craig, P., Dieppe, P., Macintyre, S., Health, P., Unit, S., Michie, S., Nazareth, I., and Petticrew, M. 2019, *Developing and Evaluating Complex Interventions*, Medical Research Council.

Craig, P., Dieppe, P., Macintyre, S., Mitchie, S., Nazareth, I., and Petticrew, M. 2008, "Developing and Evaluating Complex Interventions: The New Medical Research Council Guidance", *The BMJ*, 337, 7676, 979-83.

Craver, C.F. and Darden, L. 2013, *In Search of Mechanisms: Discoveries Across the Life Sciences*, Chicago: The University of Chicago Press.

Darden, L. and Craver, C.F. 2002, "Strategies in the Interfield Discovery of the Mechanism of Protein Synthesis", *Studies in History and Philosophy of Biolology & Biomedical Sciences*, 33, 1-28.

Ekhaguere, O.A., Oluwafemi, R.O., Badejoko, B., Oyeneyin, L.O., Butali, A., Lowenthal, E.D., and Steenhoff, A.P. 2019, "Automated Phone Call and Text Reminders for Childhood Immunisations (PRIMM): A Randomised Controlled Trial in Nigeria", *The BMJ Global Health*, 4, 2, 1-9.

Gardner, B., Davies, A., McAteer, J., and Michie, S. 2010, "Beliefs Underlying UK Parents' Views towards MMR Promotion Interventions: A Qualitative Study", *Psychology, Health and Medicine*, 15, 2, 220-30.

Gibson, D.G., Ochieng, B., Kagucia, E.W., Were, J., Hayford, K., Moulton, L.H., Levine, O.S., Odhiambo, F., O'Brien, K.L., and Feikin, D.R. 2017, "Mobile Phone-Delivered Reminders and Incentives to Improve Childhood Immunisation Coverage and Timeliness in Kenya (M-SIMU): A Cluster Randomised Controlled Trial", *The Lancet Global Health*, 5, 4, e428-e438.

Haji, A., Lowther, S., Ngan'Ga, Z., Gura, Z., Tabu, C., Sandhu, H., and Arvelo, W. 2016, "Reducing Routine Vaccination Dropout Rates: Evaluating Two Interventions in Three Kenyan Districts, 2014", *The BMC Public Health*, 16, 1, 1-8.

Illari, P.M.K. and Williamson, J. 2012, "What is a Mechanism? Thinking about Mechanisms across the Sciences", *European Journal for Philosophy of Science*, 2, 1, 119-35.

Jacobson Vann, J., Jacobson, R., Asafu-adjei, J., and Szilagyi, P. 2018, "Patient Reminder and Recall Interventions to Improve Immunization Rates", *Cochrane Database of Systematic Reviews*, 18, 1.

Jeong, J.M. and Cranney, J. 2009, "Motivation, Depression, and Naturalistic Time-Based Prospective Remembering", *Memory*, 17, 7, 732-41.

Kelly, M.P., Kelly, R.S., and Russo, F. 2014, "The Integration of Social, Behavioral, and Biological Mechanisms in Models of Pathogenesis", *Perspectives in Biology and Medicine*, 57, 3, 308-28.

Kelly, M.P. and Russo, F. 2018, "Causal Narratives in Public Health: The Difference Between Mechanisms of Aetiology and Mechanisms of Prevention in Non-Communicable Diseases", *Sociology of Health & Illness*, 40, 1, 82-99.

Littlecott, H.J., Moore, G.F., Moore, L., and Murphy, S. 2014, "Psychosocial Mediators of Change in Physical Activity in the Welsh National Exercise Referral Scheme: Secondary Analysis of a Randomised Controlled Trial", *International Journal of Behavioral Nutrition and Physical Activity*, 11, 1, 1-11.

MacDonald, N., Mohsni, E., Al-Mazrou, Y., Kim Andrus, J., Arora, N., Elden, S., Madrid, M.Y., Martin, R., Mahmoud Mustafa, A., Rees, H., Salisbury, D., Zhao, Q., Jones, I., Steffen, C.A., Hombach, J., O'Brien, K.L., and Cravioto, A. 2020, "Global Vaccine Action Plan Lessons Learned I: Recommendations for the Next Decade", *Vaccine*, 38, 33, 5364-71.

Machamer, P., Darden, L., and Craver, C.F. 2000, "Thinking about Mechanisms", *Philosophy of Science*, 67, 1, 1-25.

Manakongtreecheep, K. 2017, "SMS-Reminder for Vaccination in Africa: Research from Published, Unpublished and Grey Literature", *The Pan African medical journal*, 27, S3, 23.

Markland, D. and Tobin, V.J. 2010, "Need Support and Behavioural Regulations for Exercise among Exercise Referral Scheme Clients: The Mediating Role of Psychological Need Satisfaction", *Psychology of Sport and Exercise*, 11, 2, 91-99.

Merriam, S. and Behrendt, H. 2020, "Increasing Vaccine Uptake in Low- and Middle-Income Countries: Opportunities for Behavioural Insights Research", *Behavioural Insights Team*, https://www.bi.team/wp-content/uploads/2020/04/Opportunities-for-behavioural-insights-research-on-vaccines-uptake-in-low-and-middle-income-countries.pdf

Moore, G., Cambon, L., Michie, S., Arwidson, P., Ninot, G., Ferron, C., Potvin, L., Kellou, N., Charlesworth, J., Alla, F., Blaise, P., Bonell, C., Boutron, I., Campbell, R., Carrieri, P., Chauvin, F., Dabis, F., Edwards, N., Guevel, M.R., Kivits, J., Lacouture, A., Lang, T., Minary, L., Nour, K., Pommier, J., and Thabane, L. 2019, "Population Health Intervention Research: The Place of Theories", *Trials*, 20, 1,1-6.

Moore, G.F. and Evans, R.E. 2017, "What Theory, for Whom and in Which Context? Reflections on the Application of Theory in the Development and Evaluation of Complex Population Health Interventions", *SSM - Population Health*, 3, 132-35.

Moore, G.F., Raisanen, L., Moore, L., Din, N.U., and Murphy, S. 2013, "Mixed-Method Process Evaluation of the Welsh National Exercise Referral Scheme", *Health Education* 113, 6, 476-501.

Murphy, S.M., Edwards, R.T., Williams, N., Raisanen, L., Moore, G., Linck, P., Hounsome, N., Ud Din, N., and Moore, L. 2012, "An Evaluation of the Effectiveness and Cost Effectiveness of the National Exercise Referral Scheme in Wales, UK: A Randomised Controlled Trial of a Public Health Policy Initiative", *Journal of Epidemiology and Community Health*, 66, 8, 745-53.

Parkkinen, V.P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M.P., Norell, C., and Williamson, J. 2018, *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*, Dordrecht: Springer.

Ravallion, M. 2020, "Should the Randomistas (Continue to) Rule?", in Bédécarrats, F., Guéin, I., and Roubaud, F. (eds.), *Randomized Control Trials in the Field of Development: A Critical Perspective*, Oxford: Oxford University Press, 47-78.

Rendell, P.G., Phillips, L.H., Henry, J.D., Brumby-Rendell, T., de la Piedad Garcia, X., Altgassen, M., and Kliegel, M. 2011, "Prospective Memory, Emotional Valence and Ageing", *Cognition and Emotion*, 25, 5, 916-25.

Russo, F. and Williamson, J. 2007, "Interpreting Causality in the Health Sciences", *International Studies in the Philosophy of Science*, 21, 2, 157-70.

Sadaf, A., Richards, J.L., Glanz, J., Salmon, D.A., and Omer, S.B. 2013, "A Systematic Review of Interventions for Reducing Parental Vaccine Refusal and Vaccine Hesitancy", *Vaccine*, 31, 40 4293-4304.

Schnitzspahn, K.M., Thorley, C., Phillips, L., Voigt, B., Threadgold, E., Hammond, E. R., Mustafa, B., and Kliegel, M. 2014, "Mood Impairs Time-Based Prospective Memory in Young but Not Older Adults: The Mediating Role of Attentional Control", *Psychology and Aging*, 29, 2, 264-70.

Steel, D. 2008, *Across the Boundaries, Extrapolation in Biology and Social Science*, Oxford: Oxford University Press.

Stubbs, B.W., Panozzo, C.A., Moss, J.L., Reiter, P.L., Whitesell, D.H., and Brewer, N.T. 2014, "Evaluation of an Intervention Providing HPV Vaccine in Schools", *American Journal of Health Behavior*, 38, 1, 92-102.

Szilagyi, P.G., Albertin, C., Humiston, S.G., Rand, C.M., Schaffer, S., Brill, H., Stankaitis, J., Yo, B.K., Blumkin, A., and Stokley, S. 2013, "A Randomized Trial of the Effect of Centralized Reminder/Recall on Immunizations and Preventive Care Visits for Adolescents", *Academy of Pediatrics*, 13, 3, 204-13.

UNICEF 2018, *Levels & Trends in Child Mortality: Report 2018*.

Viner, R.M., Russell, S.J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., and Booy, R. 2020, "School Closure and Management Practices during Coronavirus Outbreaks Including COVID-19: A Rapid Systematic Review", *Lancet Child and Adolescent Health*, 4, 397-404.

WHO 2013, *Global Vaccine Action Plan 2011-2020*.

WHO 2020, *Statement—Behavioural Insights are Valuable to Inform the Planning of Appropriate Pandemic Response Measures*, https://www.euro.who.int/en/mediacentre/sections/statements/2020/statement-behavioural-insights-arevaluable-to-inform-the-planning-of-appropriate-pandemic-responsemeasures (Accessed 2020-06-29).

WHO Europe 2013, *The Guide to Tailoring Immunization Programmes (TIP) and Child Vaccination in the WHO European Region*.

Williamson, J. 2019, "Establishing Causal Claims in Medicine", *International Studies in the Philosophy of Science*, 32, 2, 33-61.

Williamson, J. 2021, "Establishing the Teratogenicity of Zika and Evaluating Causal Criteria", *Synthese*, 198, 10, 2505-18.

Worrall, J. 2007, "Why There's No Cause to Randomize", *British Journal for the Philosophy of Science*, 58, 451-88.

The Strange Numbers of Covid-19

Annibale Biggeri and Andrea Saltelli***

** University of Firenze and University of Padova*
*** University of Bergen*

Abstract

Never as with the present pandemics, numbers and the attendant activities of measuring and modelling have taken centre-stage. Yet these numbers, often delivered by academicians and media alike with extraordinary precision, rely on a rich repertoire of assumptions, including forms of bias, that can significantly skew both the numbers per se and the trust we repose in them. We discuss the issue in relation to a particular case relative to the numbers on excess mortality during the first wave of the Covid-19 pandemic in Italy. We conclude with some considerations about the use of science at the science policy interface in situations where facts are uncertain, stakes high, values in dispute and decision urgent.

Keywords: Sensitivity auditing, Sensitivity analysis, Mathematical modelling, Epidemiology, Reproducibility, Post-normal science.

1. Introduction

Increasingly, we live immersed in numbers. Numbers possess their own reactivity (Espeland and Stevens 2008); they shape the real; they are performative, seductive (Engle Merry 2016), they generate paths for new numbers to be produced in a reinforcing feedback loop (Engle Merry 2016). At a deeper level, they "create the environment that justifies their assumptions" (O'Neil 2016: 29), and endow these who produce them with legitimacy (Porter 1995).

With COVID-19, this process has received a powerful acceleration, inter-alia throwing into the limelight one mode of production of numbers—mathematical modelling, till yesterday confined to the expert communities of developers and users. With COVID-19, modelling jargon, such as "flattening the curve" (Gross and Padilla 2020), has entered into public life. For some authors, the pandemic is operating a "domestication of modelling" (Montgomery and Engelmann 2020) whereby "COVID-19 is coming to be known in maths and models" (Rhodes et al. 2020: 253):

> With COVID-19, we see that maths and models have agency as drivers of social action, translating models into citizen science and advocacy. #FlattenTheCurve

DOI 10.14275/2465-2334/202113.big

> entangles science into social practices, calculations into materialisations, abstracts into affects, and models into society (Rhodes et al. 2020: 256).

With new power, come new conflicts. "Wild-Ass Covid numbers", cries Rush Limbaugh, who adds "The minute I hear anybody start talking about models and modeling, I blanch" (Pielke 2020). Models, already suspected to be the cloak used to hide normative visions (Romer 2015), have thus become even more politicized. If the eighteen century was the century we tamed probability (Hacking 1990) (but did we?), the twenty-first is could be remembered as the century we tamed mathematical modelling (but shall we?).

Did these model produce the right COVID-19 numbers? Here the opinions are divided. John P. Ioannidis and Nassim N. Taleb, while clashing on how to interpret the pandemic, agree on one thing: the failure in model-based forecasting (Pinson and Makridakis 2020). For Caduff (Caduff 2020), one needs to interpret the model based reaction to the pandemic keeping in mind politicians' need to hide the systematic disinvestments in public health promoted by neoliberal policies - a famous report of the OECD (Organisation for Economic Co-operation and Development) warned in 2015 against excess hospital beds (OECD 2015). For Caduff, this reaction has fed into nervous media reporting, authoritarian longings and

> mathematical disease modelling—a flexible and highly adaptable tool for prediction, mixing calculations with speculations, often based on codes that are kept secret and assumptions that are difficult to scrutinize from the outside (Caduff 2020: 481).

Caduff (2020) points to the model of Imperial College (Ferguson et al. 2020) as responsible for having triggered a chain reaction of—in his view excessive—responses. We can add to Caduff that the model of the Imperial College was held responsible of "Jarring the U.S. and the U.K. to Action" (Landler and Castle 2020), while many complained about the non-easy accessibility/readability of the Imperial College model, made of thousands of lines of coding without much by way of comments. This model is agent-based, and hence intrinsically stochastic. Also for this, its reproducibility was questioned. According to the same Ferguson (2020), experts from GitHub and Microsoft supported his team check the code, which was then made available.

As per reproducibility, some researchers from Edinburgh University found a discrepancy of almost 80.000 deaths, even when using the same inputs and pseudo-random numbers for the Monte Carlo simulations. Apparently this was not due to the stochastic nature of the model, but to a bug in the code (see the discussion at https://bit.ly/2TknWR7).

Additionally, the journal Nature commented that the revised version of the program was approved by Codecheck, a project which awards "certificate of executable computation"(Eglen 2020; Singh Chawla 2020). Nature adds that other scientists have verified that the code is reproducible (Singh Chawla 2020). In the view of many, models such as the one used by Imperial College fall into the family of 'Chameleon models'. A chameleon model:

> asserts that it has implications for policy, but when challenged about the reasonableness of its assumptions and its connection with the real world, it changes its

> colour and retreats to being a simply a theoretical (bookshelf) model that has diplomatic immunity when it comes to questioning its assumptions (Pfleiderer 2020: 85-86).

For a review of modelling issues in the age of COVID-19 see Saltelli et al. 2020b.

More reflexive approaches to model validation have been put forward (Eker et al. 2018; Funtowicz and Ravetz 1994; Ravetz 2003), which invite looking at a fuller set of attributes of modelling, seen as a process. This includes investigating the interests, the framings, and the expectations of the developers (Stirling 2010). A reflexive approach hunts for tacit assumptions, implicit biases in the use of a particular modelling approach, and instances of over-interpretation of the results. An example is offered by sensitivity auditing (Saltelli et al. 2013), and another by the use of model pedigrees as in NUSAP (Funtowicz and Ravetz, 1990a; van der Sluijs et al., 2005). These approaches see mathematical modelling as a tool among many to be used in the context of a participatory approach to the analysis of policies. Participation is achieved via what is called an 'extended peer community'. This formulation is due to Post-normal science (PNS), an approach to using science when facts are uncertain, decisions urgent, stakes high and values in dispute (Funtowicz and Ravetz 1994, 1993, 1990b).

In the present pandemic, the numbers of deaths and infections have taken centre-stage. For Emmanuel Didier (Didier 2020), when a set of numbers establishes itself, other possible numbers and stories are obscured. He worries that the erosion of civil liberties associated to the fight to the pandemic may thus come to be reckoned with too late, if at all.

For practitioners trained in the use of numbers, the precision of some model-based forecasts is staggering. Using cost-benefit analysis and the controversial concept of value of a statistical life, some authors conclude that social distancing in the US will lead to a net benefit of about $5.2 trillion (Thunstrom et al. 2020). While modellers are trained to understand—one would hope—the conditionality of model-based inference, model predictions are offered at face value (Saltelli et al. 2020a).

The complexities of the disease, and its rapid transformations, have also hit on another sore-point of the science-policy interface: the existing science reproducibility crisis (Ioannidis 2005; Saltelli 2018; Saltelli and Funtowicz 2017).

While the crisis is blamed by many, inter alia, on a publish or perish imperative (Edwards and Roy 2017), the pandemic accelerated by the need to produce results rapidly, and hence the dangers of non-reproducibility and the need to retract faulty papers. Studies on chloroquine and hydroxychloroquine published in the Lancet and NEJM were retracted by their own authors, when these were forced to admit that they had no access to privately owned data sources (Joseph 2020).

Scholars trained in the tradition of post normal science (Funtowicz and Ravetz 1993) wonder of this situation might eventually favour the emergence of a new model for the relation between science and society (Waltner-Toews et al. 2020). Such a model, or 'new covenant', would embrace PNS concept of an extended peer community, made of accredited experts as well as lay public, whistleblowers, investigative journalists, and the community at large. With the pandemic, the entire planet becomes on such community, "as the appropriate behaviour and attitudes of individuals and masses become crucial for a successful response to the virus" (Waltner-Toews et al. 2020). For (Waltner-Toews et al. 2020) this extended peer community is the opposite of a technocratic, number and model-based decision strategy.

Yet the present war on vaccines offers an interesting paradox. While the authorities still subscribe to a top down, broadcast model of science communication, the other side—that variously known as conspirationists or complotists, or vaccine-hesitants, appear to constitute precisely the extended kind of peer communities advocated for by PNS (Dotto et al. 2020). The democratization of expertise in the age of internet may happen to be blind to the quality of the expertise itself. For a discussion of the downside of such a democratization see also (Mirowski 2020).

Going back to the role of models, some warn against instances of modelling hubris. Here the dimension of a model is mistaken for its quality, and one observes a lack (or misapplication) of principles and practices for good model development (Saltelli et al. 2020a).

While numbers are produced ever more assertively by the responsible authorities and experts via a top down approach (Dotto et al. 2020), increased scepticism is manifested by the general public.

What went wrong? Is the public right to look at evidence-based policy with suspicion (Saltelli and Giampietro 2017), a fortiori when based—or including steps of—mathematical models (Saltelli and Funtowicz 2014)? Are the numbers of COVID-19 indeed 'strange'?

2. A Case Study

To help answering we present a brief review of the numbers on excess mortality during the first wave of the Covid-19 pandemic in Italy. Excess mortality is defined as the difference between the total number of deaths—all-causes mortality—and the expected number of deaths—i.e. the counterfactual number of deaths it would have been observed in absence of pandemic. It has been advocated as a better measure of impact because it is less affected by reporting bias. (Beaney et al. 2020) Indeed, the reporting of Covid-19 as cause of death depends on not uniquely defined coding rules, and this bias was more severe during the first wave of the pandemic. On international comparisons, this reporting bias was considered one of the reasons for the mortality gradient between countries—for example the higher counts for Italy (Pearce et al. 2020).

Excess mortality is an interesting indicator also because it permits to take into account indirect effects of the pandemic, a disease burden consequent to the overall societal response to the pandemic. Positive and negative effects could eventually sum up when considering the total number of deaths.

This appealing indicator however depends strongly on the calculation of the expected death counts, a modelling exercise with its inherent assumptions.

On June 2020, we retrieve 5 studies evaluating excess mortality during the first wave of Covid-19 in Italy (Table 1). On May 2020, the official statistics (National Institute of Health and Italian Statistical Institute) reported a Covid-19 death counts around 35,000ths.

Scortichini et al. presented an analysis of the excess mortality across the Italian provinces, stratified by sex, age group and period of the outbreak (Scortichini et al. 2020). The analysis was performed using a two-stage interrupted time-series design using daily mortality data for the period January 2015-May 2020. In the first stage, they performed province-level quasi-Poisson regression models, with smooth functions to define a baseline risk while accounting for trends and weather conditions and to flexibly estimate the variation in excess risk during the outbreak. Estimates were pooled among provinces in the second stage using a mixed-effects multivariate meta-analysis. In the period 15 February-15 May 2020, they estimated an excess of 47,490 [95% empirical confidence intervals (eCIs): 43,984 to 50,362] deaths in Italy.

	Period	Unit of analysis	Excess deaths (95% Confidence/Credibility Interval)	Method	Age stratification
Scortichini 2020	15/2- 15/5	Province	47,490 (43,984-50,362)	Time-series	YES
Blangiardo 2020	1/1- 31/4	Municipality	41,030 (35,600-42,099)	Space-time Bayesian	NO
Alicandro 2020	29/2- 31/5	All country	46,000 (uncertainty intervals not available)	Descriptive statistics	NO
Modi 2020	16/2- 9/5	Municipality	49,000- 53,000 (only estimates from two different modelling assumptions)	Synthetic Control	YES
Magnani 2020	1/3-15/4	Municipality	45,033 (42,761-45884)	Mortality rates	YES

Table 1. Italian studies on national Covid-19 excess mortality estimates (with uncertainty intervals) available in the literature in June 2020.

Blangiardo et al. (2020) used all-cause mortality weekly rates by municipality, based on the first four months of 2016-2019. They modelled municipality weekly trends for 2016-2019, separately for males and females, for each week and year using a Poisson distribution with age-adjusted expected number of cases as offset. They specified a Bayesian hierarchical model on the log mortality relative risk allowing a province-specific temporal trend. Finally, they included a smoothed effect on weekly temperature at municipality level. They then used the output from the model for 2016-2019 to predict the number of deaths expected for each week of the 2020 follow-up period. They estimated 41,030 excess deaths (95% credibility interval CrI 35,600;42,099).

Alicandro et al. (2020) used the number of deaths registered in the first six months of 2020 and compared it with that registered in the previous quinquennium. There was an over 50% excess total mortality in March and a 38% excess in April, corresponding to over 46,000 excess deaths in those two months.

Modi et al. (2020) performed a counterfactual time series analysis on 2020 mortality data from towns in Italy using data from the previous five years as control. Specifically, they constructed a counterfactual for every region, i.e. the expected mortality counts under the scenario that the pandemic had not occurred. They then compare this counterfactual with the reported total mortality numbers for 2020 to obtain an excess death rate. They estimated that the number of COVID-19 deaths in Italy is between 49,000 and 53,000 as of May 9 2020.

Magnani et al. (2020) analyzed data by region, sex and age, and compared to expected from 2015-2019. The reference daily mortality rates were computed dividing the 2015-2019 average by the corresponding population. Ninety-five percent confidence intervals (95% CI) were computed assuming a Poisson distribution of the observed deaths [21]. Five-day moving averages were used to reduce random variation in the graphical presentation. Within the municipalities studied, 77,339 deaths were observed in the period between March 1st to April 15th, 2020, in contrast to the 50,822.6 expected. The extrapolation to the total Italian population suggests an excess of 45,033 deaths in the study period.

Almost all studies used the five years 2015-2019—one study limits to 2016-2019—as comparison to estimate the counterfactual expected mortality. The methods varied ranging from times-series to Bayesian spatio-temporal, a few based on simple averaging among previous years death counts. What is surprising is the overall agreement, considering the lower-upper limits of the estimates the variation was 35,600-53,000. All estimates oscillated around the crude difference of 46,000 between the deaths counts on 2020 and the average death counts 2015-2019.

Do these calculations represent a clear cut evidence of excess mortality or they reflect a partially conscious adherence to a given narrative of the pandemic, confirming the reluctance of investigators to depart from previous results and existing narratives (Blastland et al. 2020)?

Blastland et al. (2020) argued that scientists were more prone to persuade than to inform. We would rather change to "scientists were willing to inform on what they were persuaded". To clarify this aspect we report a study performed by Biggeri et al. (2020) on excess mortality in Italy. They averaged by season and noticed that there was a reduction in mortality in Italy during January and February 2020 compared to the previous five years. Indeed, in winter 2019-2020 there was no influenza epidemic and therefore the population later exposed to the Covid-19 pandemic was more fragile than the average population of the previous five years. The amount of susceptible people was larger because it was not harvested during a milder winter. The naïve comparison with the same months of the previous years was then biased, because the populations were not comparable—the 2020 population being more frail. Biggeri et al. obtained, for each municipality, the posterior predictive distribution under a hierarchical null model. This allowed to take into account the natural variability of the phenomenon and avoid the rigid comparison of the observed number of deaths to a fixed number of expected counts. They calculated 25,700 (95%CrI 15,963; 51,045) excess deaths for the two months of March and April 2020. The position of the authors was to assume that small variations around the expected value of mortality should be considered natural and not be counted as excess mortality.

It is of interest to notice that the large part of results on excess mortality in the literature at that time were consistent with the upper limit of the credibility interval reported in Biggeri et al. The point is that there is an underreporting of the uncertainty in the statistical estimates of excess mortality and we provide an example from Italy. This attitude of the scientists reflects the partially conscious adoption of a pessimistic point of view. This is highlighted in our Italian example by the position of most estimates closed to the upper limit of the Biggeri study.

The difference between the point estimate of the Biggeri study and the others point estimates in the literature is what is called structural uncertainty or model uncertainty, which should be added to the sampling variability which is summarized by the confidence interval.

Moreover, if we consider a trade-off between false discovery and false non-discovery, most of the results in the literature correspond to a pessimistic figure about the pandemic. The false discoveries are the number of results falsely declared positive—i.e. the number of municipalities we declared to have experienced an excess mortality while this would have not been true—and the false non-discoveries are the number of results falsely declared negative—i.e. the number of municipalities we declared to have not experienced an excess mortality while this would have been true. Having decided a rule for rejection of the null hypothesis of no excess mortality, we automatically fix the number of false discoveries and non-discoveries. The trade-off depends on the fact that if we select a rule that limits the number of false discoveries we will pay a larger number of non-discoveries. The larger the figure of excess deaths the larger the number of false discoveries—i.e. we are concerned of not incurring in a larger rate of false non-discovery. A pessimistic position corresponds to a strong concern about a false non-discovery.

3. Discussion

Our Italian example shows that impact estimates varied due to different methodological choices.

It just seems that it is never ‘just the data’. As noted in a recent work “observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty” (Breznau et al. 2021). Political scientists Giandomenico Majone offered and interesting question—which could also work as a motivation for sensitivity analysis: “Are the results from a particular model more sensitive to changes in the model and the methods used to estimate its parameters, or to changes in the data?” (Majone 1989: 62).

In other words, the data ‘speak’, but so do the many hypotheses and choices made by the different investigators to obtain their confession.

A similar recent example concerned the Rt index. Even here a team of UK investigators found confidence interval which did not overlap (Scientific Advisory Group for Emergencies 2020). A comment by Spiegelhalter et al. (2021) was:

> It’s good to explore the same question through competing approaches. Many independent teams come up with different estimates of the reproduction number R, from which a committee has to come to a consensus. We return to George Box’s quote: “All models are wrong; the practical question is how wrong do they have to be to not be useful”. No model will be “correct”, and the quoted uncertainty interval […] should be taken with a pinch of salt, as it assumes the model is the truth.

4. Conclusions

The present work points to the opportunity for more reflexive approaches, both in the specific exercise of the art of modelling and in the broader topic of the use of evidence for policy making.

As per modelling work, it would appear that overall Covid-19 statistics seem to rely on a pessimistic assumptions (Blastland et al. 2020).[1] There is a tendency in research to reproduce previous results without exploring impact of underlying assumptions (Bailey 2017).

Where to look for a solution? For some observers, institutions, regulators and health authorities should attend to the suggestions made by PNS practitioners of new model for the relation between science and society (Waltner-Toews et al. 2020). As discussed, this approach should replace the top-down science communication model presently adopted by authorities, and in a sense parallel the strategies deployed by the skeptics themselves (Dotto et al. 2020). One can note that the present crisis of expertise in the age of internet parallels what happened because of the introduction of the movable-types press in the XV century. The printing machine brought to a drastic reduction of the power of the Church and to a long season of religious wars, ending in the second half of the XVII century. It is still early to say where the present crisis will lead. Seen in this context, the case of COVID-19 and of its conflicted numbers is perhaps a small episode, but it is instructive nevertheless.

References

Alicandro, G. et al. 2020, "COVID-19 Pandemic and Total Mortality in the First Six Months of 2020 in Italy", *Med Lav*, 111, 351-53, doi: 10.23749/mdl.v111i5.10786.

Bailey, D.C. 2017, "Not Normal: The Uncertainties of Scientific Measurements", *R. Soc. Open Sci.*, 4, 160600, doi: 10.1098/rsos.160600.

Beaney, T. et al. 2020, "Excess Mortality: The Gold Standard in Measuring the Impact of COVID-19 Worldwide?", *Journal of the Royal Society of Medicine*, 113, 329-34, doi: 10.1177/0141076820956802.

Biggeri, A. et al. 2020, "A Municipality-Level Analysis of Excess Mortality in Italy in the Period January-April 2020", *Epidemiol. Prev*., 44, 297-306, doi: 10.19191/EP20.5-6.S2.130.

Blangiardo, M. et al. 2020, "Estimating Weekly Excess Mortality at Sub-national Level in Italy during the COVID-19 Pandemic", *PLOS ONE*, 15, 10, e0240286, doi: 0.1371/journal.pone.0240286.

Blastland, M. et al. 2020, "Five Rules for Evidence Communication", *Nature*, 587, 362-64, doi: 10.1038/d41586-020-03189-1.

Breznau, N. et al. 2021, "Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty", doi: 10.31222/osf.io/cd5j9.

[1] The study of Biggeri et al. 2020 does not provide the "most correct" estimate. There is not a most correct model (Saltelli et al. 2014). Pessimistic/Optimistic are relative to each other, the study of Biggeri et al. can be labelled as optimistic. We comment on the pessimistic assumption because most of the estimates tends to be higher comparing to those from Biggeri et al. 2020.

Caduff, C. 2020, "What Went Wrong: Corona and the World after the Full Stop", *Medical Anthropology Quarterly,* 34, 4, 467-87.

Didier, E. 2020, "Politique du nombre de morts", *AOC Anal. Opin. Crit.*, https://aoc.media/opinion/2020/04/15/politique-du-nombre-de-morts/.

Dotto, C. et al. 2020, "The "Broadcast" Model no longer Works in an Era of Disinformation" [WWW Document], *First Draft,* https://firstdraftnews.org:443/latest/the-broadcast-model-no-longer-works-in-an-era-of-disinformation/ (accessed 4.30.21).

Edwards, M.A. and Roy, S. 2017, "Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition", *Environ. Eng. Sci.*, 34, 51-61, doi: 10.1089/ees.2016.0223.

Eglen, S.J. 2020, "CODECHECK Certificate 2020-010", [WWW Document], *Zenodo*, doi: 10.5281/zenodo.386 5491.

Eker, S. et al. 2018, "Practice and Perspectives in the Validation of Resource Management Models", *Nat. Commun.* 9, 5359, doi: 10.1038/s41467-018-07811-9.

Engle Merry, S. 2016, *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking,* Chicago: University of Chicago Press.

Espeland, W.N. and Stevens, M.L. 2008, "A Sociology of Quantification", *European Journal of Sociol*ogy, 49, 401-36, doi: 10.1017/S0003975609000150.

Ferguson, N.M. 2020, "Microsoft and GitHub Working with the Imperial College", [WWW Document], *Twitter*, https://twitter.com/neil_ferguson/status/124183545 6947519492.

Ferguson, N.M. et al. 2020, "Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand", [WWW Document], *Imperial College London*, https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf.

Funtowicz, S. and Ravetz, J.R. 1994, "The Worth of a Songbird: Ecological Economics as a Post-normal Science", *Ecological Economics,* 10, 197-207, doi: 10.1016/0921-8009(94)90108-2.

Funtowicz, S. and Ravetz, J.R. 1993, "Science for the Post-Normal Age", *Futures*, 25, 739-55, doi: 10.1016/0016-3287(93)90022-L.

Funtowicz, S. and Ravetz, J.R. 1990a, *Uncertainty and Quality in Science for Policy*, Dordrecht: Kluwer.

Funtowicz, S. and Ravetz, J.R. 1990b, "Post-Normal Science: A New Science for New Times", *Scientific European*, 169, 20-22.

Gross, J. and Padilla, M. 2020, "Coronavirus Glossary: Flattening the Curve, Pandemic, Covid-19 and More ", *The New York Times*, March 18 2020, https://www.nytimes.com/2020/03/18/us/coronavirus-terms-glossary.html.

Hacking, I. 1990, *The Taming of Chance*, Cambridge: Cambridge University Press.

Ioannidis, J.P.A. 2005, "Why Most Published Research Findings Are False", *PLOS Medicine*, 2, 8, 696-701, doi: 10.1371/journal.pmed.0020124.

Joseph, A. 2020, "Lancet, NEJM Retract Covid-19 Studies that Sparked Backlash", *Statnews*, https://www.statnews.com/2020/06/04/lancet-retracts-major-covid-19-paper-that-raised-safety-concerns-about-malaria-drugs/.

Landler, M. and Castle, S. 2020, "Behind the Virus Report That Jarred the U.S. and the U.K. to Action", *The New York Times*, April 2.

Magnani, C. et al. 2020, "How Large Was the Mortality Increase Directly and Indirectly Caused by the COVID-19 Epidemic? An Analysis on All-Causes Mortality Data in Italy", *Int. J. Environ. Res. Public. Health*, 17, 3452, doi: 10.3390/ijerph17103452.

Majone, G. 1989, *Evidence, Argument, and Persuasion in the Policy Process*, New Haven: Yale University Press.

Mirowski, P. 2020, "Democracy, Expertise and the Post-Truth Era: An Inquiry into the Contemporary Politics of STS", *Academia.edu*, https://www.academia.edu/42682483/Democracy_Expertise_and_the_Post_Truth_Era_An_Inquiry_into_the_Contemporary_Politics_of_STS.

Modi, C. et al. 2020, "How Deadly is COVID-19? A Rigorous Analysis of Excess Mortality and Age-dependent Fatality Rates in Italy", *medRxiv*, doi: 10.1101/2020.04.15.20067074.

Montgomery, C. and Engelmann, L. 2020, "Epidemiological Publics? On the Domestication of Modelling in the Era of COVID-19", [WWW Document], http://somatosphere.net/2020/epidemiological-publics-on-the-domestication-of-modelling-in-the-era-of-covid-19.html/ (accessed 5.15.20).

OECD (Organisation for Economic Co-operation and Development) 2015, *Fiscal Sustainability of Health Systems Bridging Health and Finance Perspectives*, Paris: OECD Publishing, doi: 10.1787/9789264233386-en.

O'Neil, C. 2016, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Random House.

Pearce, N. et al. 2020, "Accurate Statistics on COVID-19 Are Essential for Policy Guidance and Decisions", *Am J Public Health*, 110, 7, 949-51.

Pfleiderer, P. 2020, "Chameleons: The Misuse of Theoretical Models in Finance and Economics", *Economica*, 87, 81-107, doi: 10.1111/ecca.12295.

Pielke, R. Jr. 2020, "The Mudfight Over 'Wild-Ass' Covid Numbers Is Pathological", *Wired*, https://www.wired.com/story/the-mudfight-over-wild-ass-covid-numbers-is-pathological/.

Pinson, P. and Makridakis, S. 2020, "COVID-19: Ioannidis vs. Taleb", [WWW Document], *Int. Inst. Forecas*t, https://forecasters.org/blog/2020/06/14/covid-19-ioannidis-vs-taleb/ (accessed 6.17.20).

Porter, T.M. 1995, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton: Princeton University Press.

Ravetz, J.R. 2003, "Models as Metaphors", in Kasemir, B., Jäger, J., Jaeger, C.G. and Gardner, M.T. (eds.), *Public Participation in Sustainability Science: A Handbook,* Cambridge: Cambridge University Press.

Rhodes, T. et al. 2020, "A Model Society: Maths, Models and Expertise in Viral Outbreaks", *Crit. Public Health*, 30, 253-56, doi: 10.1080/09581596.2020.1748310.

Romer, P. 2015, "Mathiness in the Theory of Economic Growth", *Am. Econ. Rev*, 105, 89-93, doi: 10.1257/aer.p20151066.

Saltelli, A. 2018, "Why Science's Crisis Should not Become a Political Battling Ground", *Futures*, 104, 85-90.

Saltelli, A. et al. 2013, "What Do I Make of your Latinorum? Sensitivity Auditing of Mathematical Modelling", I*nt. J. Foresight Innov. Policy*, 9, 213-34, doi: 10.1504/IJFIP.2013.058610.

Saltelli, A. et al. 2020a, "Five Ways to Ensure that Models Serve Society: A Manifesto", *Nature*, 582, 482-84, doi: 10.1038/d41586-020-01812-9.

Saltelli, A. et al. 2020b, "Five Ways to Make Models Serve Society: A Manifesto – Supplementary online material", *Nature*, 582.

Saltelli, A. and Funtowicz, S. 2017, "What Is Science's Crisis Really about?", *Futures*, 91, 5-11.

Saltelli, A. and Funtowicz, S. 2014, "When All Models Are Wrong", I*ssues Sci. Technol.,* 30, 79-85.

Saltelli, A. and Giampietro, M. 2017, "What Is Wrong with Evidence Based Policy, and How Can it Be Improved?", *Futures*, 91, 62-71, doi: 10.1016/j.futures.2016.11.012.

Scientific Advisory Group for Emergencies 2020, "SPI-M-O: Consensus Statement on COVID-19", [WWW Document], GOV.UK., https://www.gov.uk/government/publications/spi-m-o-consensus-statement-on-covid-19-8-october-2020 (accessed 5.19.21).

Scortichini, M. et al. 2020, "Excess Mortality during the COVID-19 Outbreak in Italy: A Two-Stage Interrupted Time-Series Analysis", *Int. J. Epidemiol.*, 49, 1909-17, doi: 10.1093/ije/dyaa169.

Singh Chawla, D. 2020, "Critiqued Coronavirus Simulation Gets Thumbs up from Code-checking Efforts", *Nature*, 582, 323-24, doi: 10.1038/d41586-020-01685-y.

Spiegelhalter, D. and Masters, A. 2021, "Covid Vaccines Saved Lives in England, but Why Do Estimates Differ?", *The Guardian*, July 4, https://www.theguardian.com/theobserver/commentisfree/2021/jul/04/covid-vaccines-saved-lives-england-but-why-do-estimates-differ.

Stirling, A. 2010, "Keep it Complex", *Nature*, 468, 1029-31, doi: 10.1038/4681029a.

Thunstrom, L. et al. 2020, "The Benefits and Costs of Flattening the Curve for COVID-19", *SSRN Electron. J.,* doi: 10.2139/ssrn.3561934.

van der Sluijs, J.P. et al. 2005, "Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System", *Risk Anal.*, 25, 481-92, doi: 10.1111/j.1539-6924.2005.00604.x.

Waltner-Toews, D. et al. 2020, "Post-Normal Pandemics: Why COVID-19 Requires a New Approach to Science", [WWW Document], *STEPS Cent. Blog*, https://steps-centre.org/blog/postnormal-pandemics-why-covid-19-requires-a-new-approach-to-science/

The Immunity Capital

Paolo Vineis,[*] Monica Di Fiore,[**]
Tommaso Portaluri,[***] Andrea Saltelli[****]

[*] *Imperial College*
[**] *National Research Council, Rome*
[***] *IN Srl, Udine, and Centre for Excellence and Transdisciplinary Studies, Turin*
[****] *University of Bergen*

Abstract

This paper is inspired by a thesis on "immune capital" by Kathryn Olivarius. We suggest that the biological capital, which immunity capital is part of, should be considered as an additional component of the life-course experience of individuals, together with the traditional Bourdieu's social, economic and cultural capitals that drive their lives. Building upon this concept, we consider the relationships between science, society and policy-making in the course of the pandemic. We suggest that we need to 'reframe problems so that their ethical dimensions are brought to light' (Jasanoff), with a request for humility extended to political leaders, to 'look beyond science' in search for ethical solutions. The present pandemic plays out—and is integral to—the acceleration of the rate of change, Pope Francis' peculiar word "rapidification", i.e. a vortex involving technoscience, policy and the new media.

Keywords: COVID-19, Pandemics, Embodiment, Rapidification, Media.

1. COVID-19: Nomothetic or Idiographic?

COVID-19 has accelerated our understanding of how science works and how it relates to political decision making. From a methodological point of view, we can consider the history of the epidemic in the light of the (probably obsolete) dichotomy between *nomothetic* and *idiographic* disciplines. Consider the—still largely fragmentary—causal reconstruction of the origins of COVID-19, or the issues related to immunity: it is very difficult to recognise "covering laws" here, like those valid in other fields of biology, while we more often have to resort to narratives. The latter include randomized experiments on the effectiveness of vaccines, longitudinal follow-up studies, smaller investigations on the immunological response to the virus or to vaccines, etc. There is no single modality to establish causality in such narratives. Indeed, let us consider the reconstruction involving relationships with wilderness, trade in live animals, bats as coronavirus reservoirs, intermediate animals such as pangolins, and finally the spillover to humans. In this

Argumenta 7, 1 (2021): 109-116
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.vin

circumstantial chain, the key element is the identification of a very similar sequence between the RNA of the bat-hosted virus, that of the pangolin and that of the human virus. The reference to a covering law lies in the theory of evolution in its neo-Darwinian version, in which evolution occurs by mutation and selection. But everything else in the narrative is "idiographic"—that is, local, historical, contextual, circumstantial—though some recurring patterns are identifiable. Such narratives can be used only partially for prediction; some are stronger, such as randomized trials or well-designed longitudinal studies demonstrating the effectiveness of vaccines, but others (particularly on the origin of the disease) can be used for future prediction in an indirect and incomplete way.

Let's consider immunity. What is nomothetic about it? We suggest the general principles of immunology, which nevertheless constitute a theoretical framework of reference rather than a "covering-law". Here, too, the prevailing element is narrative, proceeding by trial and error: we can measure immunoglobulins (IgM, IgG, IgA, etc.), but we do not know exactly how long immunity lasts or what degree of protection it provides. What is even more important than "humoral" immunity (from circulating antibodies) is the other form of immunity, the cellular one linked to T lymphocytes. Everything we know originates from cumulative systematic observations and repeated attempts, not really from covering laws as the philosophy of science has theorised in the past for other branches of sciences.

To explain what we mean by the interaction of the covering laws (nomothetic) and the idiographic components we can refer to the great medical historian Mirko Grmek, who coined the term "pathocenosis" to describe the relationship between human diseases and the surrounding environment in its historical determinations (Grmek 1969). These relationships are largely based on the role of the immune system and the combination of mutation and selection (and this is the nomothetic part). The fact that an infectious disease can cause a pandemic depends on two main conditions: the state of immunological susceptibility to infection of the entire population (a condition that occurs, for example, when a virus is new to humans) and the aetiological agent's ability to transmit itself efficiently from one person to another (and this is the idiographic, i.e. circumstantial, part). The emergence of a new virus or bacterium is not a rare occurrence, but is part of normal evolutionary processes. The ability to infect new animal species, including humans, gives these mutated viruses or bacteria a selective advantage as they are able to expand into new ecological niches. The impact of a pandemic on the population depends on the spread rate, the severity of the clinical picture, and the lethality rate. The spread rate is measured by a basic reproduction index (indicated by R_0), i.e. the average number of people infected (secondary cases), in a population that is fully susceptible to every single contagious case. The index is directly proportional to the duration of contagiousness of the infected person and the frequency of contacts during which other people are exposed to the infection. It is therefore understandable why social distancing measures are effective by reducing one of the parameters of R_0. Several infectious agents, even common ones, have a very high reproduction index under conditions of complete susceptibility of the population. For example, for measles it is usually estimated that an average of 12-18 secondary cases occur for each primary case (Guerra, Bolotin, Lim, Heffernan, Deeks, Li, Crowcroft 2017). However, for most circulating infections, the majority of the population has now developed immunity, both because people have already contracted the infection or because they have been vaccinated against it. The proportion of people immune to a specific infection in the population is a hindrance to the spread of the infection, as every immune person—even if exposed to infection—is supposed not to infect anyone else. When the proportion of immune persons in a population is so high that it does not allow for epidemic spread, it is said that the population has developed "herd immunity" (a questionable term: humans are not a herd, that is they have a much more complex social structure, mobility, and autonomy than animals). Most of the determinants and characteristics of the spread of the infection in a population seem to be circumstantial, non-deterministic and local.

The severity of the clinical picture is another important parameter regarding the relevance of the pandemic and its spread. Infections characterised by a severe clinical picture are more easily detected and, if the contagiousness is limited to the period of time in which the symptoms occur, they are easier to intercept. These aspects may explain why some infections cause pandemics and others only epidemics, which remain circumscribed. The SARS epidemic of 2003, for example, was more contained than the current spread of COVID-19: the R_0 was similar, but the mortality rate was significantly higher (9-16%), and above all, the highest infection rate was found in the second week after the onset of symptoms. These characteristics facilitated the identification of cases and the isolation of people exposed to the infection (their contacts), leading to the eradication of the epidemic before it spread broadly such as COVID-19. Once again, this was circumstantial rather than depending on nomothetic explanations.

Indeed, every epidemic has its own characteristics, linked to the type of aetiological agent, the harm it induces and the way it is transmitted, so it is difficult to derive countermeasures from one epidemic to another and predict their course by analogy. The behaviour of different pathogens can be estimated by constructing mathematical models that simulate the conditions of infection transmission, produce possible spread scenarios and offer the possibility of evaluating the effect of specific countermeasures (Saltelli, Bammer, Bruno, Charters, Di Fiore, Didier, Nelson Espeland, Kay, Lo Piano, Mayo, Pielke, Portaluri, Porter, Puy, Rafols, Ravetz, Reinert, Sarewitz, Stark, Stirling, van der Sluijs, Vineis 2020). So the history of the spread of each epidemic or pandemic has a nomothetic component, in the sense that it can be interpreted in the light of the virus' mutations, its adaptation to the host and the latter's immune response. But the overall narrative has many more "idiographic" elements, linked to chance (for example the appearance of the right mutation at the right time), and to the geographical and historical context. The term *pathocenosis* encompasses these aspects as well as the unstable balance between different diseases in a population; the pathocenosis is constantly in a condition of precarious balance or imbalance. It should be noted that the human (or animal) response to microbial aggression also leads to a constant instability of the immune system, which is subject to continuous recombinations of its genetic material for the production of antibodies corresponding to new environmental antigens.

2. Immunocapital

The term immunity has acquired metaphorical meanings in addition to the scientific meaning. Communities respond to threats through responses that Roberto Esposito called "immune", playing on the link between "immunitas/communitas" (latin for immunity/community) (Esposito 2015). Immune reactions, according to Esposito, are a common way in which human communities respond to external and internal threats: not only disease, but also economic crises, migration and so on. In fact, Esposito's metaphor is not only a social and philosophical concept, but has become a reality in certain historical periods as we suggest below.

An eloquent case of an intertwining of a scientific and a social use of the word immunity is that of yellow fever in New Orleans, as described in Kathryn Olivarius' PhD thesis at Oxford University (Kofler and Baylis 2020, Olivarius 2016). For much of the nineteenth century, the inhabitants of New Orleans were divided between those who were "acclimated" to the risk of yellow fever and everyone else. The former could marry, work and even, if they were slaves, alter their market value. Many young people organised "parties" in order to become infected and increase their market value, especially among immigrants. That choice was much more dangerous than today, because yellow fever was a terrible disease, with very serious symptoms and a 50% lethality rate.

According to Olivarius, yellow fever gave rise to a new social stratification, and in this sense the researcher coined the term "immuno-capital", which adds to the categories proposed by Pierre Bourdieu: economic, social and cultural capital. In this same direction, within the European network "Lifepath" (Vineis et al. 2020), some of us proposed a fourth type of capital, the biological one, which has not yet been given enough thought. What Bourdieu omits, in fact, is precisely the biological component, even if it is partly contained in his idea of "habitus" (consisting of bodily aspects such as posture and accent, and more abstract mental characteristics such as patterns of perception and classification).

Bourdieu does not fully consider the ways in which the three types of capital he describes all have—in different and synergistic ways—an impact on the body and health condition of the subject. Income, culture and social capital, both separately and together, are capable of changing life paths towards better or worse health. There is indeed a mutual relationship between the three types of capital and "biological capital": a congenital (e.g. cognitive) birth defect can severely affect access to culture, social capital and income; and, conversely, a low capital of each of the three types strongly influences the subject's state of health in its biographical development. Social class influences the posture, the demeanour, the physical appearance of the person, and this in turn "condemns" them to a certain class membership (which was especially true in the past, when for example the "low" classes in England were also physically shorter than the "high" classes). Our proposal (Vineis et al. 2020) is therefore that together with biography (*bios*) we consider, in close interaction with it, the biology (*zoe*) of people.

The biological capital includes the immune capital. To return to the New Orleans example, in 60 years over 150,000 people died in 22 epidemic waves in the capital of Louisiana. Insurers were reluctant to cover those who were "not acclimated" (a phenomenon that has not yet occurred for COVID-19). Other typical aspects of American society at the time were accentuated by the epidemic: for example, slave traders claimed that slavery had the advantage of "distancing" rich whites from black slaves, even though the disease was transmitted by mosquitoes (but exposure to mosquitoes was much more frequent in slums). White people could stay at home, slaves moved around to work in the fields and were therefore more exposed to disease—another analogy with today's situation. These marked class differences, according to Olivarius, meant that at that time those in power had little interest in implementing prevention and containment measures. Even today, social differences are marked by COVID-19 infection and mortality rates (although much less so than then): in the British OpenSafely study, COVID-19 mortality rates were found to be more than twice as high in people living in more "deprived" areas compared to rich areas (Williamson et al. 2020).

3. Science and Politics

Let us examine more in depth the complex relationship between science and political decision making. An emblematic case is that of the recent "immunity licences", based on serological tests, which became famous for a very short period of time in 2020 (at least in Italy) and then proved to be impracticable. The hope was that the measurement of immunoglobulins could release from isolation, allow people to continue working, the elderly to feel protected, and everyone to go on holiday. But the sensitivity of the tests was low, the antibodies measured were not "neutralising" against the virus and the temporal relationships with the clinical history of the disease were very uncertain. So no licence. Beyond the scientific inconsistency of the proposal (and the practical problems that came with it: it was not possible to test the entire population at a given time), the immunity licence is an example of the many amplifications of pre-existing problems that emerged with COVID-19. These include social inequalities, access to treatment, the right to health, ethical dilemmas (do we save everyone? protect the elderly? protect the economy?), the conflict between small businesses and multinationals, national selfishness towards solidarity, and top-down interventions vs. individual responsibility. More recently the introduction of the "green pass" has raised similar discussions, though it is a completely different case compared with the immunity licences: the green pass is an instrument—based on nudging—to obtain that people have access to essential services and a normal life by inducing the majority to get vaccinated.

It is undeniable that political power uses science and technology to avoid taking a stance on complex issues (Saltelli, Bammer, Bruno, Charters, Di Fiore, Didier, Nelson Espeland, Kay, Lo Piano, Mayo, Pielke, Portaluri, Porter, Puy, Rafols, Ravetz, Reinert, Sarewitz, Stark, Stirling, van der Sluijs, Vineis 2020). Consider the case of Rt, at the centre of press debates and political decisions—a sort of barometer of the trend of the epidemic and of the effectiveness of containment interventions (note that R_0 measures the virus' transmissibility in a completely susceptible population, Rt in a population in which at least some people have become immune). Rt has certain technical characteristics that cannot be ignored if one wants to interpret it correctly: (a) Rt is based only on symptomatic cases; (b) Rt is subject to random fluctuations if the cases are limited in number, and this should be taken into account by associating a statistical confidence interval to it. Politics and the media have largely ignored the intrinsic technical characteristics of Rt, leading to erroneous inferences.

In the spectrum of positions that characterise the relations between science and society, there are some extreme ones such as "denialism", which (despite their differences) is rooted in Romanticism or in radical thinkers like Schmitt. However, today there is a propensity to use science as a surrogate for choices that should be primarily about values, and this is the case both among institutional actors urging actions such as confinement or vaccination and among those resisting the same policies. Privileging science and technology may work fine in an emergency phase, but not in a planning phase where it is essential to explicitly refer to values and, based on these, to bring out predictive scientific models that explore different scenarios. While in the first phase of the epidemic it was understandable to rely entirely on science to find answers, and it was also justifiable to make drastic choices such as lockdowns on the basis of mathematical models—whose assumptions were not completely explicit (see Saltelli, Bammer, Bruno, Charters, Di Fiore, Didier, Nelson Espeland, Kay, Lo Piano, Mayo, Pielke, Portaluri, Porter, Puy, Rafols, Ravetz, Reinert, Sarewitz, Stark, Stirling, van der Sluijs, Vineis 2020)—in the following phases this attitude is no longer acceptable. Now it is really important to clarify the underlying values and to approach science based on those, so as to guide the scientific (reproducible and intersubjective) exploration of the various hypothetical scenarios.

Even during the emergency, in reality, lockdown measures were only necessary in a relative and conditional way, i.e. as tools needed to achieve certain types of (moral and political) goals. It is these goals that are in question in the public debate. Lockdown measures should be defined as "just" rather than necessary, for example because they have made it possible to safeguard the most fragile part of the population (essentially the elderly and the sick), keeping alive the feeling of social solidarity and the intergenerational pact. But now they must be reconsidered in the light of similar and explicit value considerations. Obviously, many questions remain open and should be at the centre of the public debate. At what levels should values enter the debate and the decisions be related to public health? What if disagreement about values occurs? How might trade-offs be established, and who should establish them?

For example, the inversion of the relationship between ethics-politics and science could consist in this: formulating some policy-making scenarios and asking researchers to quantify their consequences, including economic ones. The scenarios could be: (a) a Kantian scenario in which not even a single life is sacrificed (as far as possible); (b) a utilitarian one which calculates the greatest benefit and the least damage for the greatest number of people; (c) a weighted utilitarian one, which gives more importance to the lives of young people, etc. For each of these, it would be up to the modellers to assess the implications in terms of lives lost, intensive therapies, economic degrowth, prospects for future generations, and so on. Note that the political style adopted by Trump, Bolsonaro and to a lesser extent Johnson corresponds to yet another scenario, the ultraliberal individualistic one. Another obvious problem in terms of values, rather than technical issues, is *who* will have access to vaccines (in the face of the tendency of rich countries to hoard them for themselves), which introduces the dimension of equity in political decisions.

It would also appear that the science policy system has not yet metabolized the long list of surprises and front reversal brought about by the pandemic, where the winning and losing countries exchange place with surprising rapidity—all phenomena largely unpredicted by the existing apparatus of prediction and control. The same apparatus that in recent years has become more apt to influence electoral outcomes rather than to predict pandemics: in spite of its expanding technologies, it should perhaps engage in different technologies, those of humility (Tverberg 2021). Jasanoff warns against hubris technologies, such as risk and cost-benefit analysis, that 'show peripheral blindness towards uncertainty and ambiguity'. For her, 'predictive technologies are limited in their capacity to internalise challenges that arise outside their framing assumptions' (Jasanoff 2003). Therefore, the 'binary thinking that frames the future in terms of certain choices between options knowable', cannot deliver us the entire picture and all the answers.

We can raise our awareness of the complexity by acting with *humility*, that is induce a reflexion on what we ignore, and what is uncertain, in order to 'reframe problems so that their ethical dimensions are brought to light'. Jasanoff invites to reflect on vulnerabilities, on winners and losers, and on learning opportunities. A request for humility extended to political leaders, to 'look beyond science' in searching for ethical solutions. The present pandemic plays out—and is integral to—the acceleration of the rate of change, Pope Francis' peculiar word "rapidification", or a vortex involving technoscience, policy and the new media (Pope Francis 2015, Saltelli, Boulanger 2019).

Perhaps the present pandemic has altered our pathocenosis in one important respect: that of the relation between science and policy (Waltner-Toews, Biggeri, De Marchi, Funtowicz, Giampietro, O'Connor, Ravetz, Saltelli, van der Sluijs 2020). Ruling elites can no longer rely on experts for persuading the public that their policies are beneficial, correct, inevitable, and safe. For David Waltner Toews, we have learned that not a single model nor a single policy bears all the solutions, but many models and many policies. The idea of human, animals and viruses as part of a larger set of nested hierarchies enter into collision with previous Cartesian narratives of man as master and owner of nature. The wonders of Cartesian science give us vaccines developed at an unprecedented rate; yet the world is not made of things surrounding us, but of the set of relationships holding all these together (Waltner-Toews, 2020). Will this realization impact our pathocenosis?

References

Esposito, R. 2015, *Immunitas*, Turin: Einaudi.

Grmek, M. 1969, "Préliminaires d'une Étude Historique des Maladies", *Ann. E.S.C.*, 24, 1473-83.

Guerra, F.M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S.L., Li, Y., Crowcroft, N.S. 2017, "The Basic Reproduction Number (R0) of Measles: A Systematic Review", *Lancet Infect Dis.*, 17, e420-e428.

Jasanoff, S. 2003, "Technologies of Humility: Citizen Participation in Governing Science", *Minerva*, 41, 223-44.

Kofler, N. and Baylis, F. 2020, "Ten Reasons Why Immunity Passports Are a Bad Idea", *Nature*, 581, 379-81.

Olivarius, K. 2016, "Necropolis. Yellow Fever, Immunity and Capitalism in the Deep South, 1800-1860", PhD thesis, Oxford, https://ora.ox.ac.uk/objects/uuid:749d8bac-63a7-4afe-9696-5cfa40dfc854.

Pope Francis 2015, *Laudato Si'*, Vatican City: Libreria Editrice Vaticana.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke, R. Jr, Portaluri, T., Porter, T.M., Puy, A., Rafols, I., Ravetz, J.R., Reinert, E., Sarewitz, D., Stark, P.B., Stirling, A., van der Sluijs, J., Vineis, P. 2020, "Five Ways to Ensure That Models Serve Society: A Manifesto", *Nature*, 582, 482-84.

Saltelli, A., Boulanger, P., 2019, "Technoscience, Policy and the New Media: Nexus or Vortex?", *Futures*, 115, 102491, doi: 10.1016/J.FUTURES.2019.102491.

Tverberg, G. 2021, "We Can't Expect COVID-19 to Go Away: We Should Plan Accordingly", *Our Finite World*, https://ourfiniteworld.com/2021/04/11/we-cant-expect-covid-19-to-go-away-we-should-plan-accordingly/, retrieved April 11.

Vineis, P. et al. 2020, "Special Report: The Biology of Inequalities in Health: The Lifepath Consortium", *Front Public Health*, 118, 1-37.

Waltner-Toews, D. 2020, *On Pandemics: Deadly Diseases from Bubonic Plague to Coronavirus*, new edition, Vancouver: Greystone Books.

Waltner-Toews, D., Biggeri, A., De Marchi, B., Funtowicz, S., Giampietro, M., O'Connor, M., Ravetz, J., Saltelli, A., van der Sluijs, J. 2020, "Post-Normal Pandemics: Why COVID-19 Requires a New Approach to Science", *STEPS Centre Blog*, March 26.

Williamson, E. et al. 2020, "Factors Associated with COVID-19-Related Death Using OpenSAFELY", *Nature*, 584, 430-36.

Making Best Use of the Available Evidence: Mechanistic Evidence and the Management of the Covid-19 Pandemic

Virginia Ghiara

University of Kent

Abstract

In this paper, I argue that evidence of biological and socio-behavioural mechanisms can contribute to the management of Covid-19. I discuss two examples that show how scientists are using different forms of evidence, among which mechanistic evidence, to answer questions about the efficacy of vaccines against Covid-19 and the effectiveness of vaccination interventions in different contexts. In the first example I claim that, due to the fast pace of the pandemic, mechanistic reasoning and evidence of biological mechanisms play an important role in the study of vaccines' efficacy and the development of new adaptations based on possible future virus mutations. In the second example, I explore the use of evidence of the socio-behavioural mechanisms influencing vaccination behaviours and I show that the World Health Organisation is promoting the collection of this type of evidence to understand whether particular vaccination interventions can fit in local contexts. Overall, this discussion supports the claim that the dominant evidence-based medicine (EBM) approach, which relies heavily on difference-making studies to assess the effectiveness of clinical and public health intervention, is inadequate and should be replaced by a new approach, EBM+, that systematically considers mechanistic studies alongside association studies.

Keywords: Causation, Evidential pluralism, Mechanistic evidence, Mechanistic reasoning.

1. Introduction

In 2007 Federica Russo and Jon Williamson introduced a version of evidential pluralism according to which:

> To establish causal claims, scientists need the mutual support of mechanisms and dependencies. […] The idea is that probabilistic evidence needs to be accounted for by an underlying mechanism before the causal claim can be established (Russo and Williamson 2007: 159).

ISSN 2465-2334

DOI 10.14275/2465-2334/202113.ghi

The account put forward by Russo and Williamson, known by the name of the Russo-Williamson thesis, challenged the dominant evidence-based medicine (EBM) approach, which relies heavily on difference-making studies to assess the effectiveness of clinical and public health interventions. According to the EBM approach, causation can be established if it is possible to establish a probabilistic relationship between the cause and the effect (the intervention A causes B only if A raises the probability of the occurrence of B), or a counterfactual relationship between them (A and B are actual events, and if A had not occurred, then B would not have occurred).

Among difference-making studies, randomised controlled trials (RCTs) have been often described as 'the gold standard', whereas other types of studies are often not considered as useful when establishing if a clinical or public health intervention is effective. In particular, mechanistic studies that examine the mechanism through which an intervention exerts its impact on the effect, are often disregarded as less useful (Guyatt et al. 1992). In several versions of EBM, in fact, mechanistic knowledge is partly taken into account, and probabilistic correlations are based on some forms of causal models (see for instance Howick 2011, and Spirtes et al 2000). This form of knowledge, however, is not given the same relevance that is given to difference-making studies.

The thesis suggested by Russo and Williamson paved the way for the EBM+ approach, a development of EBM that treats mechanistic studies on a par with association difference-making studies. Based on the more nuanced picture of causal assessment proposed by the Russo-Williamson thesis, since in real-world studies it is often impossible to completely rule out the possibility that a difference-making relationship is due to bias, confounders, or relationships other than causation, it is important to include an explicit scrutiny of mechanistic studies when assessing causal claims. What distinguishes causal from spurious correlations is the presence of a mechanism between A and B, therefore causation can be established only if it is possible to establish the existence of a mechanism of action as well as the existence of a correlation.[1]

The EBM+ approach carefully distinguishes between types of evidence and types of evidence-gathering methods: difference-making studies mostly generate direct evidence that the putative cause A and the putative effect B are correlated, but they can also provide evidence that indirectly supports the existence of a mechanism (see for instance Illari 2011). Mechanistic studies, on the other hand, can provide not only mechanistic evidence, but also evidence of a correlation between A and B. In most cases, however, establishing the presence of a mechanism requires direct evidence from mechanistic studies, which helps to confirm or disconfirm mechanistic hypotheses.

In a recent paper, Aronson et al. (2020) reviewed the role of mechanistic reasoning in four major areas that are relevant to the management of Covid-19: treatments, pharmacological surveillance, preventative public health interventions and vaccination programmes. Aronson et al. published their article when vaccines against Covid-19 were still under development. Over the last 6 months, however,

[1] The Russo-Williamson thesis has generated interest both in the health and in the social sciences. The thesis has not been immune to critiques, and some authors have discussed counter-examples to the Russo-Williamson (see for instance Claveau 2012, Klement and Bandyopadhyay 2019, Reiss 2009). Some of these criticisms have been examined in Ghiara 2019.

the debate on Covid-19 has been dominated by discussions about the real efficacy of the current vaccines against variants, and current vaccination behaviours. Due to the coronavirus pandemic's pace, the opportunities to conduct RCTs to explore such issues are still limited, and there are situations when scientists and policymakers have combined different sources and types of evidence to understand how to best manage Covid-19. In July 2021, a systematic review of randomized controlled trials assessing the effect and safety of COVID-19 vaccines identified only 14 trials assessing 10 types of COVID-19 vaccines, the majority of which on phase I or II (Chen et al. 2021). Emani et al. (2020), moreover, analysed RCTs that assessed possible treatment options and concluded that that literature was very limited and most studies were characterised by significant methodological limitations.

In this paper, I will review the role that mechanistic evidence is playing in addressing such challenges.

2. Causal Mechanisms and Mechanistic Evidence

The term mechanism can be understood broadly in three ways: i) as a complex system consisting of entities and activities organised in such a way that, together, they are responsible for the phenomenon under study (as described by Machamer et al. 2000); ii) as a mechanistic process through time and space; this process can be understood as a process propagating a signal (Dowe 2007; Reichenbach 1958; Salmon 1997) or as a chain of events that leads to specific effects (as described in the social sciences by Maxwell 2004), and iii) as a combination of a complex system and a process.

Although there are differences between biological and social mechanisms, this categorisation can be applied to different types of mechanisms. The complex molecular system of long-term memory, organised in entities (neurons, proteins and genes) and activities (such as protein movements and gene expressions), and Schelling's well-known social segregation mechanism (1978), organised in individuals and their discriminatory preferences are two examples of complex systems consisting of organised entities and activities. The propagation of an electrical signal from an artificial peacemaker to the appropriate part of the heart, and the political causal chains leading to revolutions identified by Skocpol (1979), moreover, are examples of causal processes (for more on this point, see Ghiara 2019).

Mechanistic evidence, in turn, is evidence that supports the existence of a mechanism.

3. Efficacy and Effectiveness

When examining the use of mechanistic reasoning, it is important to distinguish between efficacy and effectiveness. The term 'efficacy' refers to the effect of some intervention in ideal conditions. Establishing efficacy is typically the first step to evaluate whether an intervention works. For instance, in the case of Covid-19 vaccines, establishing efficacy would require evidence that the vaccine can reduce Covid-19 incidence under optimal conditions within a study population.

Study populations, however, often differ from the target population in significant ways. For example, a study population for evaluating Covid-19 vaccines might exclude those with multiple morbidities or young people; or a study population for evaluating vaccination campaigns might exclude minority groups. For this reason, establishing whether an intervention actually works requires investigating its 'effectiveness', which means the effect of the intervention 'in the real world', in the target population. Evidence of mechanisms plays a crucial role in establishing both efficacy and effectiveness (Parkkinen et al. 2018). The sections below discuss two examples of how mechanistic evidence is being used to establish vaccines' efficacy and vaccination programmes' effectiveness.

4. Assessing the Efficacy of Vaccines Against Current and Future Covid-19 Variants

It is well known that mutations are a normal part of viruses' life cycles, and that through such mutations new variants are likely to arise. Multiple Covid-19 variants have been documented in the last 6 months, and their discovery has been followed by long debates on whether such variants will persist, and what impact these can have on the efficacy of the current Covid-19 vaccines.

Randomised controlled trials appear to support the claim that most of the existing variants will not completely undercut the efficacy of this first generation of vaccines (Madhi et al. 2021; Abdool Karim and de Oliveira 2021),[2] however concerns have emerged over variant 501Y.V2 found in South Africa and Brazil, also known as variant B.1.351, as new evidence showed that it was able to evade virus-blocking antibodies produced by most people previously infected with first-wave strains, and that some of the existing vaccines have a reduced efficacy against it (Callaway and Ledford 2021).

The questions scientists need to answer are not only whether current vaccines are able to protect against current variants, but also whether they will be able to protect against future variants of the virus. Answering these questions is, undoubtedly, very challenging. As mentioned by Mascola et al. (2021), this situation is like tackling a moving target. On the one hand, there is limited time to evaluate vaccines' efficacy against current variants, on the other hand answering the second question requires making predictions about the key features of future variants. In this situation, mechanistic evidence can help both to address the first question and to develop new hypotheses about what adaptations or other types of vaccines might work against future variants.

A case in point showing how mechanistic evidence can contribute to the response against the virus' variants is the recent study published by Clark et al. (2021). In their study, the authors reported results from experiments conducted on lab-made, non-infectious reproductions of eight Covid-19 mutations found in a patient receiving immune-suppressive treatments for an autoimmune disorder. Over 5 months, Clark et al. reported, these mutations had clustered on a section of the spike called the receptor-binding domain. This finding attracted the researchers' attention, as the spike is what current antibody treatments and vaccines target to prevent Covid-19 from entering human cells. Through a series of experiments the authors showed that, of the eight mutations reproduced in their laboratory, two evaded both antibodies naturally occurred in people who survived the infection, and lab-made antibodies now used for the clinical treatment of Covid-19.

[2] My argument refers to the first generation of vaccines, which include vaccines based on different technologies including mRNA vaccines, inactivated vaccines, viral vector vaccines and subunit vaccines (Chen et al. 2021).

Interestingly, the mutations analysed by Clark and colleagues have not yet been identified in the mutations already detected. However, the authors highlighted the importance of this mechanistic evidence since "How the spike responded to persistent immune pressure in one person over a five-month period can teach us how the virus will mutate if it continues to spread across the globe" (Pesheva 2021). Thinking about the mechanisms of mutation, in turn, can help to determine whether the existing vaccines would work against such mutations, and what adaptations would be successful against them.

In particular, this evidence has advanced some mechanistic hypotheses on how a next generation of vaccines should target less mutable segments of the virus to work against mutations. Scientists have started examining T cells, immune cells that can target and destroy virus infected cells, to look for evidence suggesting that such cells could help to preserve lasting immunity. The hypothesis that T cells could play a role in providing immunity partly relies on mechanistic reasoning and background knowledge: as reported by Daina Graybosch, a biotechnology analyst at investment bank SVB Leerink in New York City, although data is still not sufficient to draw conclusion, "it makes sense biologically" (Pesheva 2021). Researchers have focused their attention on two groups of T cells. The first group is that of killer T cells (or CD8 T cells), which identify and destroy cells that are infected with the virus. Another group of T cells, called helper T cells (or CD4 T cells), support the production of antibodies and killer T cells. Based on these cells' functions, researchers hope that T cells could destroy the virus-infected cells before they spread from the upper respiratory tract, and could reduce transmission by reducing the amount of virus circulating in an infected person. Additional evidence on T cells, furthermore, suggests that they could be less vulnerable to Covid-19 mutations: Sette et. al (2021), examined that infected people generally produce T cells targeting at least 15-20 different portions of coronavirus proteins.

Mechanistic hypotheses, of course, can offer only partial support to understand vaccinations' efficacy against current and future variants, and what adaptations might be required. In such uncertain times, however, this evidence helps to identify potential pathways and future research directions.

5. Examining Barriers and Enablers and the Effectiveness of Vaccination Programmes

As claimed in the "Tailoring Immunization Programmes" guide published by the WHO's Regional Office for Europe (2019), it is crucial to understand the psychological, contextual, and social mechanisms that influence vaccination behaviours in order to design an effective campaign. In October 2020, the WHO advisory group reviewed some of the mechanistic evidence concerning possible barriers to vaccination and published the report "Behavioural considerations for acceptance and uptake of COVID-19 vaccines". Through a review of the literature, the authors identified three categories of barriers and enablers in relation to vaccine uptake: environment, social influences, and motivation.

Social influence was recognised as a main factor when promoting vaccination against Covid-19, and the authors recommended several supporting strategies based on mechanistic studies. For instance, it was observed that the general beliefs of a community and the corresponding behaviours are likely to influence the individual attitude towards vaccination, and that if vaccine uptake is made "visible" to others (e.g. through social signalling such as badges or via social media), members of community will more easily perceive vaccine uptake as consistent with their community's social norm (Karing 2018; WHO 2020). It was also considered that the behaviours of respected members of the community, who share similar values and characteristics with the targeted group (for instance, with the same religious or ethnic identity), are likely to influence vaccination uptake within their community (CASS 2020).

Motivation to get vaccinated is also linked, according to the report, to the perceived risk of contracting Covid-19, or to the perceived health consequences of the infection. Behavioural studies showed that most people use shortcuts to assess risks in complex circumstances, and their perception might be based on personal experiences and rumours (Kahneman 1973; Tversky et al. 1974). It follows that clear communication is crucial in order to help people judge risks accurately. A behavioural mechanism that appears to influence perceived risks and motivation is known by the name of "anticipated regret": when people anticipate that a negative outcome in the future would lead them to wish they had behaved differently (Brewer et al. 2016; Brown et al. 2010). This mechanism can be used in favour of vaccination, suggested the authors, through the description of the consequences of not getting vaccinated (for instance health practitioners might discuss with patients how they would feel if they do not get vaccinated and get infected or transmit the virus to their family).

As predictable, most of the studies used to develop such recommendations had been published before the current pandemic, and were focused on specific geographical areas, or specific communities. For instance, the social signalling recommendation was mainly based on a study conducted in Sierra Leone, while the role played by members of the community was explored through studies of Sub-Saharan African communities. Moreover, most of the studies were focused on other health concerns (such as seasonal flu or Ebola), and some targeted only child vaccination and parents' behaviours. It follows that it is questionable whether health policies based on such recommendations would be effective to promote vaccinations against Covid-19 in different contexts.[3]

To answer this question, policymakers need to use their knowledge that something had a particular impact somewhere to extrapolate that the same thing will exert a similar impact in a different context. Going from evidence of efficacy ('this has a given effect in an experimental population in context X') to evidence of effectiveness in a different context ('this will have the same effect in a target population in context Y') can be challenging. For instance, Nancy Cartwright (2010) showed that untested assumptions about the possibility of extrapolating evidence from a nutrition program conducted in Southern India to Bangladesh led to a big failure due to substantial differences in the social contexts.

The risk of falling into the same trap when developing Covid-19 vaccination strategies has not been discussed in detail in the report published by WHO in October 2020, but it is certainly one of the challenges policymakers face when using general recommendations regarding Covid-19 vaccinations. Although

[3] An additional caveat needs to be added in the case of behavioural studies: internal validity and the replicability of the experiments can be questioned too. The problem of replicability has been discussed in relation to several experiments on cognitive biases (see for instance Romero 2019, and Schimmack, Heene and Kesavan 2017). It follows that the use of evidence from behavioral studies in health policies needs to be carefully examined.

evidence shows that the virus impacts in very similar ways all populations, this does not mean that the same barriers and enablers, and the corresponding vaccination strategies, will work in the same way in different contexts.

Understanding the behavioural and social mechanisms operating in different contexts, hence, is an important step to ensure vaccination behaviours are promoted in an effective way. This consideration is one of the reasons why WHO published, in February 2021, a new guide entitled "Data for action: achieving high uptake of COVID-19 vaccines" (WHO 2021). In this guide, which was then updated in April 2021, the authors adapted the Brewer's general mechanistic model to Covid 19. The assumption is that, regardless of the context, all the included factors can play a role, however their influence might vary in different contexts, and the effectiveness of targeted strategies might change when applied to different populations.

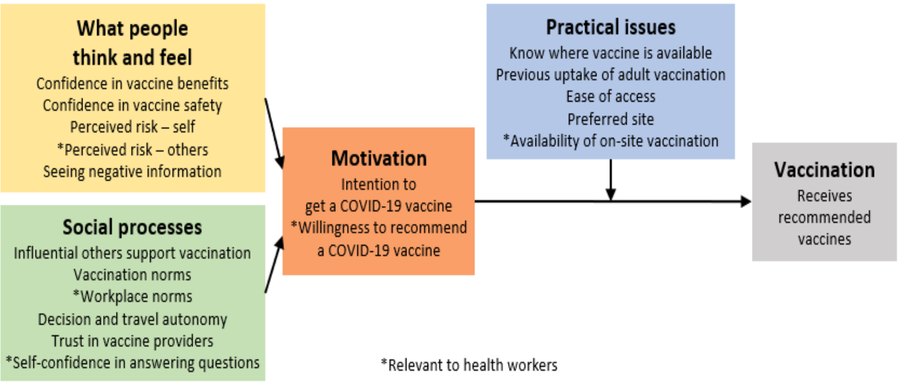

What people think and feel
Confidence in vaccine benefits
Confidence in vaccine safety
Perceived risk – self
*Perceived risk – others
Seeing negative information
Social processes
Influential others support vaccination
Vaccination norms
*Workplace norms
Decision and travel autonomy
Trust in vaccine providers
*Self-confidence in answering questions
Motivation
Intention to get a COVID-19 vaccine
*Willingness to recommend a COVID-19 vaccine
Practical issues
Know where vaccine is available
Previous uptake of adult vaccination
Ease of access
Preferred site
*Availability of on-site vaccination
Vaccination
Receives recommended vaccines
*Relevant to health workers

Figure 1: The adaptation of Brewer's model that explores personal, social, motivational and practical barriers to Covid-19 vaccinations (WHO 2021: 16).

Using the adapted Brewer's model, a global group of experts have developed survey questions and qualitative interview topic guides to help national policymakers collect mechanistic evidence and identify how such factors impact vaccination in a given context. Why is this process important? Understanding the local mechanisms of actions is a critical aspect to test whether the barriers and enabling mechanisms in the target population are sufficiently similar to those in the study population where the study or experiment was conducted. This, in other words, helps to investigate the external validity of particular causal links between environment, social influences, and motivation, and vaccination behaviours.

Understanding such similarities between the study and the target populations, in turn, can help policymaker to understand whether an "intervention is feasible in specific contexts" (WHO 2021: 16) or, put differently, whether it is possible to expect that an intervention (such as on-site vaccination or an educational campaign) will work equally well in the study and target population.

A case in point that illustrates how the same campaign might have different (or even opposite) effects is the example of a herd immunity campaign: knowledge of the importance of herd immunity might lead people to get vaccinated to protect others, but could also lead to free-riding behaviours, where people avoid individual costs of vaccination because they know they will benefit from others' immunisation. Some studies have identified differences between contexts: Betsch et al.

(2017), for instance, observed that emphasising social benefit appears to work better in Western cultures, that tend to be more individualistic, than in more collectivist Eastern cultures. Making the individual benefits of a herd immunity too obvious, on the other hand, might have a stronger impact on Western culture, where more people might decide to free-ride (Betsch et al. 2013).

As the example above shows, a detailed understanding of the similarities and differences between the relevant mechanisms in the study and the target populations is crucial to ensure vaccination interventions will be effective, and can complement difference-making evidence collected in the study and targeted populations.

6. Conclusion

In this paper, I illustrated how evidence of biological and socio-behavioural mechanisms can contribute to the management of Covid-19. The two cases described above are illustrative and not exhaustive, as other studies on Covid-19 are likely to use different forms of evidence to answer different questions.

The first case I discussed shows that, since the fast pace of the pandemic limits the possibility of running randomised controlled trials and requires scientists to design adaptations based on possible future virus mutations, mechanistic reasoning and evidence of biological mechanisms can play an important role to determine the current and future efficacy of vaccines against Covid-19. In the second case I explored the use of evidence of the socio-behavioural mechanisms influencing vaccination behaviours. I showed that the World Health Organisation is promoting the collection of this type of evidence to understand whether particular vaccination interventions can fit in local contexts, and I claimed that mechanistic evidence can play a crucial role to establish external validity and extrapolate interventions.

This paper does not want to argue that evidence of mechanisms is sufficient to answer causal questions concerning Covid-19. To assess the efficacy of vaccines and to establish if a vaccination intervention works in a local context, difference-making evidence is essential. The fast pace of the pandemic, however, requires scientists and policymakers to make fast decisions, and trials often require time to ensure the collection of robust difference-making evidence. As a consequence, this uncertain time casts a new light on the benefits of using mechanistic evidence and on the limitations of the traditional evidence-based medicine (EBM) approach.

References

Abdool Karim, S.S. and de Oliveira, T. 2021, “New SARS-CoV-2 Variants-Clinical, Public Health, and Vaccine Implications”, *New England Journal of Medicine*, 384, 1866-68.

Aronson, J.K., Auker-Howlett, D., Ghiara, V., Kelly, M.P., and Williamson, J. 2020, “The Use of Mechanistic Reasoning in Assessing Coronavirus Interventions”. *Journal Of Evaluation in Clinical Practice*, doi: 10.1111/jep.13438.

Brewer, N.T., DeFrank, J.T., and Gilkey, M.B. 2016, “Anticipated Regret and Health Behavior: A Meta-Analysis”. *Health Psychol*, 35, 11, 1264-75.

Brown, K.F., Kroll, J.S., Hudson, M.J., Ramsay, M., Green, J., Long, S.J., Vincent, C.A., Fraser, G., and Sevdalis, N. 2010, "Factors Underlying Parental Decisions About Combination Childhood Vaccinations Including MMR: A Systematic Review", *Vaccine*, 28, 26, 4235-48.

Callaway, E. and Ledford, H. 2021, "How to Redesign COVID Vaccines So They Protect Against Variants", *Nature*, 590, 7844, 15-16.

Cartwright, N. 2010, "Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps", *Philosophy of Science*, 79, 5, 973-89.

Cellule d'Analyse en Sciences Sociales (CASS) 2020, *Humanitarian Programme Recommendations For COVID-19 Based on Social Sciences Evidence from The DRC Ebola Outbreak Response. Social Science Support For COVID-19: Lessons Learned Brief 3*, www.unicef.org/drcongo/media/4131/file/CASS-Brief3-recommendations.pdf

Chen, M., Yuan, Y., Zhou, Y., Deng, Z., Zhao, J., Feng, F., and Sun, C. 2021, "Safety of SARS-Cov-2 Vaccines: A Systematic Review and Meta-Analysis of Randomized Controlled Trials", *Infectious Diseases of Poverty*, 10, 1, 1-12.

Clark, S.A., Clark, L.E., Pan, J., Coscia, A., McKay, L.G., Shankar, S., Johnson, R.I., Brusic, V., Choudhary, M.C., Regan, J., and Li, J.Z. 2021, "SARS-CoV-2 Evolution in an Immunocompromised Host Reveals Shared Neutralization Escape Mechanisms", *Cell*, 184, 10, 2605-17.

Claveau, F. 2012, "The Russo–Williamson Theses in The Social Sciences: Causal inference Drawing on Two Types of Evidence", *Studies in History and Philosophy of Science Part C: Studies In History And Philosophy of Biological And Biomedical Sciences*, 43, 4, 806-13.

Dowe, D. 2007, *Physical Causation*, Cambridge: Cambridge University Press.

Emani, V.R., Goswami, S., Nandanoor, D., Emani, S. R., Reddy, N.K., and Reddy, R. 2020, "Randomized Controlled Trials for COVID-19: Evaluation of Optimal Randomization Methodologies-Need for the Data Validation of The Completed Trials, and to Improve the Ongoing and Future Randomized Trial Designs", *International Journal of Antimicrobial Agents*, doi: 10.1016/j.ijantimicag.2020.106222.

Ghiara, V. 2019. *Inferring Causation from Big Data in The Social Sciences*, Doctoral dissertation, University of Kent.

Guyatt, G., Cairns, J., Churchill, D., Cook, D., Haynes, B., Hirsh, J., Irvine, J., Levine, M., Levine, M., Nishikawa, J., and Sackett, D. 1992, "Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine", *Jama*, 268, 17, 2420-25.

Howick, J.H. 2011, *The Philosophy of Evidence-Based Medicine*, Hoboken: John Wiley and Sons.

Illari, P.M. 2011, "Mechanistic Evidence: Disambiguating the Russo-Williamson Thesis", *International Studies in the Philosophy of Science*, 25, 2, 139-57.

Kahneman, D. 1973, *Attention and Effort*, Englewood Cliffs: Prentice-Hall.

Karing A. 2018, "*Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone*" [working paper], Berkeley: University of California.

Klement, R.J. and Bandyopadhyay, P.S. 2019, "Emergence and Evidence: A Close Look at Bunge's Philosophy Of Medicine", *Philosophies*, 4, 3, 50.

Madhi, S.A., Baillie, V., Cutland, C.L., Voysey, M., Koen, A.L., Fairlie, L., Padayachee, S.D., Dheda, K., Barnabas, S.L., Bhorat, Q.E., and Briner, C. 2021, "Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine Againstthhe B. 1.351 Variant", *New England Journal of Medicine*, doi: 10.1056/NEJMoa2102214.

Machamer, P., Darden, L., and Craver, C.F. 2000, "Thinking about Mechanisms", *Philosophy of Scienc*e, 67, 1, 1-25.

Parkkinen, V.P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M.P., Norell, C., Russo, F., Shaw, B., and Williamson, J. 2018, *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*, London: Springer Nature.

Pesheva, E. 2021, *How a Mutated Coronavirus Evades Immune System Defenses*, https://news.harvard.edu/gazette/story/2021/03/how-sars-cov-2-evades-immune-system-defenses/

Reichenbach, H. 1958, "The Direction of Time", *The Philosophical Quarterly*, 8, 30, 72.

Reiss, J. 2009, "Causation in the Social Sciences: Evidence, Inference, and Purpose", *Philosophy of the Social Sciences*, 39, 1, 20-40.

Romero, F. 2019, "Philosophy of Science and The Replicability Crisis", *Philosophy Compass*, 14, 11, doi: 10.1111/phc3.12633.

Russo, F. and Williamson, J. 2007, "Interpreting Causality in The Health Sciences", *International Studies in the Philosophy of Science*, 21, 157-70.

Salmon, W.C. 1997, "Causality and Explanation: A Reply to Two Critiques"*, Philosophy of Science,* 64, 3, 461-77.

Sette, A. and Crotty, S. 2021, "Adaptive Immunity to SARS-CoV-2 and COVID-19", *Cell*, doi: 10.1016/j.cell.2021.01.007.

Schelling, T. 1978, *Micromotives and Macrobehavior*, New York: W.W. Norton and Company.

Schimmack, U., Heene, M., and Kesavan, K. 2017, "Reconstruction of a Train Wreck: How Priming Research Went Off the Rails", *Replicability Index*.

Skocpol, T. 1979, *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*, New York and Cambridge: Cambridge University Press.

Spirtes, P., Glymour, C.N., Scheines, R., and Heckerman, D. 2000, *Causation, Prediction, and Search*, Cambridge, MA: MIT press.

Tversky A, Kahneman D. 1974, "Judgment under Uncertainty: Heuristics and Biases", *Science*, 185, 4157, 1124-31.

Tversky, A. and Kahneman, D. 1973, "Availability: A Heuristic for Judging Frequency and Probability", *Cognitive Psychology*, 5, 207-32.

World Health Organization 2021, *Data for Action: Achieving High Uptake Of COVID-19 Vaccines*.

World Health Organization 2020, *Behavioural Considerations for Acceptance and Uptake Of COVID-19 Vaccines: WHO Technical Advisory Group on Behavioural Insights and Sciences for Health*.

World Health Organization 2019, *The Guide to Tailoring Immunization Programmes (TIP): Increasing Coverage of Infant and Child Vaccination in The WHO European Region*.

Monitoring the Safety of Medicines and Vaccines in Times of Pandemic: Practical, Conceptual, and Ethical Challenges in Pharmacovigilance

Elena Rocca and Birgitta Grundmark***

* *Centre for Applied Philosophy of Science, Norwegian University of Life Sciences*
** *Department of Surgical Sciences, Uppsala University*

Abstract

In this paper, we analyse some of the challenges that pharmacovigilance, the science of detecting and assessing possible adverse reactions from medical interventions, is facing during the COVID-19 pandemic. In particular, we consider the issue of increased uncertainty of the evidence and the issue of dealing with an unprecedented amount of data. After presenting the technical advances implemented in response to these two challenges, we offer some conceptual reflections around such practical changes. We argue that the COVID-19 emergency represents a chance to push forward critical thinking in the field of pharmacovigilance, and that contributions from epistemology, ethics and philosophy of science are necessary to increase resilience in the face of this and future health emergencies.

Keywords: COVID-19, Pharmacovigilance, Resilience, Big Data.

1. Introduction

The coronavirus disease (COVID-19) pandemic emerged in Wuhan, China, in 2019 and rapidly spread globally during 2020. COVID-19 is not only a crisis for public health and healthcare. It is also a challenge for the established structures of knowledge production, use and communication (Meng 2020). The COVID-19 crisis is forcing us to improve the way we make science-based decisions in the face of uncertainty. This is necessary in order to increase resilience for this and future pandemics or other health emergencies (Leonelli 2021).

In this paper, we argue that the COVID-19 emergency represents a chance to push forward critical thinking in the field of risk assessment of medical interventions. A health crisis requires urgent action from healthcare, but such urgency cannot come at the cost of patient safety. When a medicine enters the market, its safety is only partially known. Effects on vulnerable groups are often undetected within

Argumenta 7, 1 (2021): 127-146
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.roc

pre-marketing clinical trials; for this reason, a system of post-marketing monitoring is in place in order to identify evidence of possible side effects as early as possible. The aim of this paper is to show that improving safety of medicines and vaccines in cases of global health emergency is not only a practical, but also a conceptual challenge. As such, it should be met not only with technical improvements of existing processes, but also by incorporating contributions from epistemology, ethics and philosophy of science. The ultimate aim is a more self-critical, interdisciplinary and resilient practice of risk assessment of medicines and vaccines.

Pharmacovigilance is the science of detecting and assessing possible adverse reactions from medical interventions. Although in pharmacovigilance all types of evidence, including laboratory research, observational studies and anecdotal reports are potentially crucial, most of post-marketing safety monitoring is based on the so called 'passive surveillance'. The cornerstone of this process is the spontaneous reporting of potential adverse reactions by manufacturers, health professionals or patients. Analyses of adverse reaction reports generate hypotheses about causality between medicine or vaccines and the reported symptoms. Such hypotheses of causal connections are sometimes called 'signals'. In pharmacovigilance, a signal is defined as a hypothesis of a risk from a medicine with data and arguments that support it (Uppsala Monitoring Centre 2021a).

During the COVID-19 pandemic, the challenges and complexities of safety monitoring in pharmacovigilance have been amplified (Ferreira-da-Silva, Ribeiro-Vaz, Morato & Polónia 2021). Arguably, this extra-burden is due to three factors that have changed during the COVID-19 pandemic. These are:

- *Increased uncertainty,* as a result of the novelty of the disease and the novelty of certain medicines and vaccines, often approved with a lower level of safety evidence as compared to medicines approval in non-urgent situations.
- *Increased amount of data to be handled and processed*, mainly because of massive COVID-19 vaccination programs in place globally.
- *Increased public attention,* due to the public perception of the state of emergency and to the extensive media coverage of issues around drugs and vaccines safety.

In the literature there are recent accounts on how manufacturers and drug authorities evolved in order to face these issues (Ferreira-da-Silva et al. 2021). Notably, much of the focus of the innovation process is set, especially by manufactures, on digitalization, automatization, and the development of more sophisticated data-mining algorithms and artificial intelligence technology to increase the effectiveness of existing procedures (ICON 2020; Pharmafile 2021). This optimistic trend rightly sees technological innovations as an important part of the solution, and some manufacturers have gone so far as saying that the COVID-19 emergency offered the chance to innovate the company's pharmacovigilance procedures, which had been otherwise stagnating and conservative. Even more optimistically, media reported that

> Using smart technology to manage the […] process not only simplifies what can be laborious and time-consuming work for humans, but can also help to reassure members of the public who are concerned about the safety of newly developed drugs (Pharmafile 2021).

However, although technological improvements are an important part of the solution, it is also known that technological innovation alone will not lead to a sustainable improvement of medicines' and vaccines' risk assessment (Naidu, Sushma, Jaiswal & Asha 2020). Here, we urge that equal attention should be given to practical, technological advances *and* to the critical reflections necessary to make such advances meaningful and efficient. Only this way can the COVID-19 experience be harnessed to improve risk and safety assessment of medical interventions.

In the following, we are going to analyse in detail two of the three COVID-19 related challenges: the issue of dealing with increased uncertainty, mainly in relation to safety monitoring of COVID-19 treatments, and the issue of handling increased amount of data, mainly in relation with safety monitoring of COVID-19 vaccines. For each of the two challenges, we first outline a general description; secondly, we give an overview of the practical implemented measures so far; finally, we present the related critical reflections and indicate some conceptual advances pushed by each specific COVID-19 related challenge.

Before starting the analysis, the next section briefly introduces the process of pharmacovigilance.

2. Safety Monitoring and Risk Assessment in Pharmacovigilance

The paradigmatic case that started the modern pharmacovigilance structure, was the thalidomide disaster, where the drug used as an antiemetic during pregnancy provoked rare limbal malformations in the new-born (Dally 1998). After this, the WHO Programme for International Drug Monitoring (PIDM) started a collaboration between the drug authorities of by now 148 countries for systematic global monitoring of all medical treatments before and after being released on the market (Letourneau, Wells, Walop & Duclos 2008).

The standard procedure of the so called 'passive surveillance' in pharmacovigilance is that observations of suspected adverse effects of medicines and vaccines, collected during regular clinical use, are reported by marketing authorization holders, health professionals or the public to the national authorities of each country member of PIDM. These reports are registered into national databases and often shared, together with some reports from pre- and post-authorization clinical trials, in international databases curated by WHO (VigiBase) and other international agencies (e.g. EudraVigilance, curated by the European Medicine Agency). For this, one needs to digitally transcribe and code the reports using standardised international terminology both for medicines, vaccines and symptoms (Mugosa, Stankovic, Turkovic, & Sahman-Zaimovic 2015). A valid adverse reaction report must contain at least coded information about an identifiable reporter, an identifiable patient, a suspected adverse reaction and a suspect medicinal product (CIOMS working group VIII 2010). Only when the data are in standardised format, can they be retrieved from databases, accessed and analysed by pharmacovigilance experts to detect new possible causal relationships between reported reactions and medicines.

Typically, the knowledge accumulation about a new adverse effect follows the shape of an S curve with three phases. A first slow generation of suspicion, followed by a rapid accumulation of case reports (signal strengthening) and a final slower period of confirmation, typically including post-marketing observational studies (Meyboom, Hekster, Egberts & Gribnau 1997).

In pharmacovigilance agencies, new hypotheses of causality do not get assessed until a sufficient number of cases accumulate. The final phase of confirmation is usually based on clinical trials and/or pharmacoepidemiological research studies, which traditionally have taken long in relation to the timelines of decision makers. Often, a preliminary regulatory decision has to be taken already during the second phase of signal strengthening and possible confirmation.

For the process of hypothesis-generation, a vast spectrum of information is used: from preclinical studies, to clinical experiments, active surveillance and observational studies. However, post marketing hypothesis generation in pharmacovigilance is primarily based on passive surveillance, as described above.

Based on the information retrievable from national and global databases, pharmacovigilance experts need to assess whether the drug is likely to play a causal role for reported symptoms, or not. There are three complementary approaches to this task:

- Single case assessment: each single report goes through an independent causality assessment. There are several methods for causality assessment in the single case, however they all have some common points (Meyboom, Hekster, Egberts & Gribnau 1997). These include: (I) considerations of temporality; (II) the presence of confounders, such as illness or other drugs, which could equally well (or better) explain the symptoms; (III) evaluations of the symptoms over time (see table 1, excerpt from WHO-UMC methodology for causality assessment).
- Analysis of case series: when a series of relevant cases is collected and identified, the hypothesis of causality is assessed by verifying whether the putative effect is consistent, robust and specific through the available cases. The Bradford-Hill criteria are often used to test the causal hypothesis, and this usually implies the consideration of many different types of evidence (pre-clinical, clinical studies, safety profile of similar drugs, etc.) (Shakir & Layton 2002).
- Statistical methods: when numbers of reports/drug event combinations are too large to be individually manually analysed, statistical measures are used as a tool to detect signals. In these cases, the likelihood of a causal hypothesis is judged by the amount of reports linking the drug to the same symptom. *Disproportionality* measures calculate whether the combination drug-symptom is reported into the database more times than expected if the combination happened by pure chance (CIOMS working group VIII 2010. Once detected as disproportionate, signals may subsequently be analysed manually.

With this short introduction to the process of hypothesis generation in pharmacovigilance, we are now going to look in more details at the way this system was challenged during the pandemic.

3. Pharmacovigilance and COVID-19 Treatments: Dealing with the Challenge of Increased Uncertainty

3.1. Why Is Evidence of Adverse Effects from COVID-19 Medicines Uncertain?

One of the issues challenging pharmacovigilance during the COVID-19 pandemic is, as mentioned, increased uncertainty. What is this uncertainty due to, and why does it impact pharmacovigilance considerably?

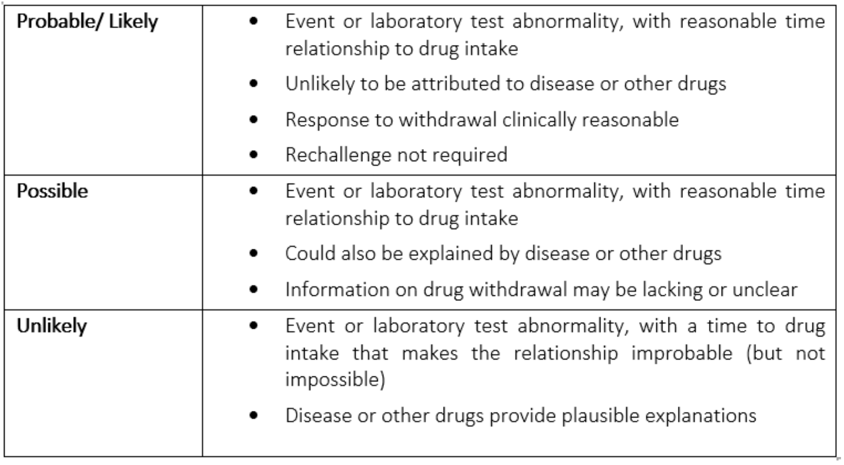
On one hand, we are dealing with a new human corona virus, SARS-CoV-2. The virus causes mild to severe pneumonia with a pathogenesis that is still to a certain extent unknown and has been gradually but still only partially elucidated during the course of the pandemic. To complicate the picture, the pathophysiology or the illness has turned out to be a particularly complex one. Respiratory distress syndrome is the primary cause of SARS-CoV-2 mortality, but the disease may affect multiple organs where heart failure, thrombo-embolic events, severe single or multiorgan dysfunction are common among causes of COVID-19 fatality (Machhi et al. 2020). It has thus been difficult, especially in the first year of the pandemic, to evaluate whether a certain reported symptom might or not be caused by the underlying COVID-19.

On the other hand, we are dealing with a health emergency in which many severely ill patients were co-medicated with a huge arsenal of medicines in the lack of an acknowledged therapeutic approach (Desai 2020). It is difficult to disentangle the causal contribution of so many medicines, given that a medicine repurposed for COVID-19 patients might have a different safety profile in this particular context. Moreover, several new treatments for COVID-19 have so far been approved for emergency use, with limited knowledge of their safety profile. COVID-19 adverse effect reports often contain a long list of co-medications, and it is difficult to evaluate whether a certain reported symptom might partially or entirely be caused by one of them (Gérard et al. 2021).

Finally, some undesired effects might be provoked by a combination of the medicine(s) used, the COVID-19 infection, and the background medical history of the patient. Indeed, risk groups for developing severe COVID-19 are weak, old and some chronically ill patients (Machhi et al 2020). At the same time, these patient groups may similarly be partly susceptible to adverse drug reactions because of declining organ functions, for instance of the liver and kidney (Mühlberg & Platt 1999). It is clinically reasonable to suppose that some of these patients may be predisposed to be hurt by a certain treatment which is otherwise safe in the majority of the population. At least in some cases, then, a certain adverse effect can be generated by the interrelation of different causal contributions in the individual patient.

To understand the extent to which this situation hinders pharmacovigilance recall that, for the causality evaluation of single adverse event reports, one decisive factor is whether the symptom can be explained by another medicine or underlying condition (see section 2). Let's consider an example. Imagine that a patient without any history of allergy and skin diseases has a rash after the initiation of an antibiotic. Say also that timing of the rash onset is compatible with the biology of the drug-body interaction, and the symptom disappears after drug cessation. According to most of the causality assessment methods (table 1), causality in this case will be categorised as 'probable' because other acknowledged causes of the event have been excluded. If, however, the patient had episodes of rash in the past, or has an infection that could explain the rash, or is already using a medicine which is associated with rash, the causality would be classified as only 'possible'.

Similarly, since it is uncertain whether a specific symptom associated with a COVID-19 treatment could be explained by the underlying COVID-19 infection, or by (a multitude of) other concomitant COVID-19 medicines, causality in the vast majority of the adverse reaction reports will at best be classified as 'possible', without further possibility of discerning among them (Desai 2020).



Probable/ Likely	• Event or laboratory test abnormality, with reasonable time relationship to drug intake • Unlikely to be attributed to disease or other drugs • Response to withdrawal clinically reasonable • Rechallenge not required
Possible	• Event or laboratory test abnormality, with reasonable time relationship to drug intake • Could also be explained by disease or other drugs • Information on drug withdrawal may be lacking or unclear
Unlikely	• Event or laboratory test abnormality, with a time to drug intake that makes the relationship improbable (but not impossible) • Disease or other drugs provide plausible explanations

Table 1. Excerpt from WHO-UMC methodology for causality assessment (Uppsala Monitoring Centre, 2021b).

3.2. How to Cope with Increased Uncertainty? Practical Implemented Measures

In this complex situation, causality assessment methods that rely on an evaluation of the difference made by each single causal factor, are of limited help. Some experts have even predicted early in 2020 that the causality assessment of single COVID-19 related reports would be impossible, and that "causation needs to be viewed for the study drug with a public health perspective" (Desai 2020).

One predominant way to face this situation has indeed been to focus on the population level, in the lack of precise single case causality assessment. This was done, for instance, for the novel antiviral remdesivir, which was granted emergency authorisation for the treatment of COVID-19 (Saint-Raymond et al. 2020). Since preclinical studies showed a potential renal toxicity, and clinical trials produced unclear results about this potential side effect, it was important to further assess the risk (Saint-Raymond et al. 2020). One explorative approach has been to search databases for the number of adverse reaction reports containing the term 'remdesivir' together with one of more terms indicating renal failure. Using a statistical disproportionality measure (called Information Component, IC), it was possible to assess that remdesivir was reported in correlation with terms of renal failure more often than expected (Gérard et al. 2021). Authors point out numerous caveats, not least the persistence of many confounding factors, nevertheless they argue that this evidence reinforces the hypothesis of harm. However, other statistical designs have reached different conclusions. For instance, a retrospective cohort study on COVID-19 patients who received remdesivir did not find a statistically significant association between the medicine and renal impairment, concluding that this particular safety warning may be a 'clinical lore' rather than a valid precaution (Ackley, McManus, Topal, Cicali, & Shah 2021). Ultimately therefore, statistical evidence is still contradictory, and whether experts adopt a cautionary mode still depends largely on their interpretation of the preclinical toxicity and pharmacology studies (Gevers, Welink, & van Nieuwkoop 2021) alongside clinical study results.

In parallel to the mainstream focus on statistical strategies to control for confounders, a second tactic was promoted by drug agencies of countries such as Norway and France (Grandvuillemin, Drici, Jonville-Bera, Micallef, & Montastruc 2021). These experts emphasise the need of efficacy and responsiveness of the system in times of health emergency and to control for confounders by improving the clinical *quality* over the quantity of the data:

> Although COVID-19 is a confounding factor per se, owing to its potential for multi-organ damage including the heart and kidney, the *quality of the transmitted data* in adverse drug reaction reports, the *timeliness of feedback from clinicians*, and the real-time pharmacological and medical analysis […] made it possible to swiftly identify relevant safety signals (Ibid: abstract, emphasis ours).

In these systems, pharmacovigilance experts use their decentralised national network of clinicians and pharmacists who contributed with detailed clinical investigations of some cases. Decentralised national pharmacovigilance systems allowed to promptly detect signals of harm for some of the COVID-19 treatments. An example is the Intracranial Venous Sinus Thrombosis, in combination with thrombocytopenia, a rare syndrome that was detected and confirmed in some individuals after immunisation with certain COVID-19 vaccines and which was quickly detected in countries such as Norway and Denmark (Norwegian Medicines Agency 2021). Moreover, the French medicines agency claimed that their system allowed early detection and communication of the cardiac adverse events occurring in some COVID-19 patients treated with hydroxychloroquine (Grandvuillemin et al. 2021). Their conclusion is that:

> Some pharmacovigilance systems are working on automated signal detection by using tools connected to very large databases. However, for the time being, these methods enable the identification of signals, but do not allow for any conclusion on a causal link, for which a medical and pharmacological evaluation remains essential. Moreover, a real-time medical and pharmacological analysis is crucial in this type of health crisis (Ibid: 407).

Clearly, 'normal business' pharmacovigilance would not see these two strategies as mutually exclusive. As explained in part 2, it is normal practice to integrate statistical and clinical approaches for causality assessment. However, it appears that in the COVID-19 emergency the role of single case assessment and clinical expertise for facing increased uncertainty is under discussion. Most experts would probably agree that in an ideal world there would be resources to both improve the sophistication of statistical studies, for instance by joining different databases and registries, *and* build up decentralised networks of clinical experts. However, resources are limited and need to be wisely allocated. Clinical causality assessment in pharmacovigilance is a resource- and time-consuming task, especially if it needs to happen in parallel with a health crisis requiring extra healthcare resources (Desai 2020). The question then becomes: is it worthy to maintain and invest resources in improving qualitative evidence of this type? Would it ultimately help building resilience to deal with future situations of increased uncertainty?

This is a practical question that hides a conceptual issue about the role of qualitative evidence for knowledge-building, and the type of scientific discoveries we seek in pharmacovigilance.

3.3 Uncertainty and Scientific Discoveries in Pharmacovigilance: A Critical Reflection

The field of pharmacovigilance is generally struggling with a tension between the need of prompt regulatory action to safeguard the health pf patients and minimize the impact of the detected adverse effects and the need of sufficiently good evidence to support the action taken, a tension that is emphasised in times of emergency. Partially, this tension is due to the low epistemic[1] role that is traditionally assigned to single case reports and qualitative evidence. There is a growing resistance against establishing causality, or expanding scientific knowledge, based on few outlier cases (Howick 2011). In the evidence-based medicine pyramid of evidence, evidence from case studies and expert opinion are rank the lowest for the purpose of establishing causality (Howick 2011). The best way of looking for causal links is generally considered controlled experimentation, where confounding factors are controlled for.

Nevertheless, the epistemic role of single case in pharmacovigilance is clearly higher than normally granted by evidence-based medicine (see part 2). The legislation states that safety warnings in the product labels should be based on "at least a reasonable possibility, based for example, on their comparative incidence in clinical trials, or on findings from epidemiological studies *and/or on an evaluation of causality from individual case reports*" (European Commission Enterprise and Industry Directorate 2009). A hypothesis of harm from a medical treatment, therefore, does not *in principle* need to be supported with statistical evidence and could be formulated on the basis of as few as three cases, or even less (ibid). Traditionally, pharmacovigilance emphasises causality assessment in the single case, and is close to a *singularist* view of causation (Uppsala Monitoring Centre 2021b). In this view, the causal link is best investigated by studying in detail the causal context and by understanding the causal processes at place (Anjum & Rocca 2019).

What, then, when the problem of confounding is major and the uncertainty is high, like in the case of the COVID-19 emergency? Should pharmacovigilance emphasise the statistical approach to try to control confounding factors, getting closer to the EBM pyramid of evidence? Or should more effort be invested in the clinical investigation of single cases, maintaining a singularist take on causation?

This question requires that we critically reflect on why pharmacovigilance has traditionally acquired such a different epistemological take on causal evidence compared to other medical disciplines.

One answer could be that pharmacovigilance is mainly an exploratory activity, which needs curiosity and "prepared minds" to identify unexpected risks (Trontell 2004). As such, it was categorised as a specific process of discovery, namely *serendipity* (Rocca, Copeland & Edwards 2019). Serendipity is the process of making a discovery when not looking for it. Serendipitous discoveries are based on the emphasis of unexpected but valuable findings (ibid). This view accurately describes the first, explorative phase of pharmacovigilance, largely based on passive surveillance and on a multitude of different types of evidence. In this sense, discoveries in pharmacovigilance are different from other discoveries in medicine, that are instead intentionally derived from an established theory (the efficacy of a

[1] In the paper, by 'epistemic' we mean 'knowledge-related'.

drug, for instance). An argument for the favourite status of the single case in pharmacovigilance is that rigid study designs are unfit for discovering the unexpected (Osimani 2013). On the contrary, successful drug safety monitoring must succeed in catching the significance of unexpected clinical observations (Rocca, Anjum & Mumford 2017). What counts for serendipitous discoveries is the quality, and not the quantity, of the observation (Copeland 2017).

A resilient pharmacovigilance system, then, would be one that promotes serendipitous discoveries, especially when a prompt reaction to crisis is needed (Rocca, Copeland & Edwards 2019). How could this be done?

There is no easy answer to this question, however some conceptual ground has been laid regarding this issue. Recent advances in serendipity research acknowledge the importance of the social context, trans-disciplinary networks, diversity of expertise and plurality of methodological perspectives (Copeland 2017). In other words, chance and the prepared mind (or sagacity, as it is also called) are not enough to catch the unexpected. An interesting observation that is not followed up by the scientific community, for instance because dismissed as "low quality" according to the dominant standards of evidence, does not lead anywhere.

Interdisciplinary responsive networks are typically formed in response to virus outbreaks. As we experienced during 2020, knowledge about the SARS-CoV-2 progresses exceptionally fast, because different disciplines collaborated closely under the perception of a common problem to solve (Leonelli 2021). In these circumstances, observations are picked up and considered by different disciplinary perspectives. Because of this, communities fighting virus outbreaks have been explicitly called "sites of serendipity" (Michener et al. 2009).

Following this reasoning, pharmacovigilance systems that emphasise decentralised network of clinical experts and encourage in-situ clinical assessment of the single cases seem in line with the promotion of a serendipitous, responsive network in which clinicians and pharmacovigilance experts collaborate with the purpose of catching unexpected clinical observations in real time. If we think in terms of serendipity, we can say that in time of pandemics the importance of informative narratives is crucial. Understanding the causal story in its contexts, including patient-generated evidence and hypotheses of inherent mechanisms at place in the specific patient, is as challenging as crucial. The French national Agency of Medicine recommendation, of keeping the clinical analysis as essential for early detection of possible side effects during the pandemic (Grandvuillemin et al. 2021), is in line with our critical reflection here.

In summary, in this session we have outlined the challenge of dealing with increased uncertainty, due to confounding from a new virus and new use of medications during the COVID-19 pandemic. We have shown that the pharmacovigilance community tried opposing strategies, from downplaying the difficult task of causality assessment in the single case in favour of a population approach, to allocating extra resources for the specific task. Finally, we have made our main point: that predicting which strategy is the most effective requires critical thinking about the specific task of pharmacovigilance and the type of evidence needed to promote it. When such considerations are made, clinical expertise and in-situ causal evaluation appear even more important when uncertainty is high.

4. Pharmacovigilance and COVID-19 Vaccines: Dealing with Big Data

4.1 Why is Big Data an Issue for Pharmacovigilance During the COVID-19 Pandemic?

As already mentioned, most of the world's countries have in place a system for the safety surveillance of medicines and vaccines on a large, population scale. In developed countries it is becoming increasingly common to base this surveillance in electronic healthcare databases, and data are often shared in common databases among countries. For instance VigiBase, the WHO global database of individual case safety reports, contains over 20 million reports of suspected adverse effects of medicines, shared, since 1968, by member countries of the WHO Programme for International Drug Monitoring (Lindquist 2008). National and international databases are periodically analysed with data-mining approaches. Such analyses may be more or less systematic depending on the mandate of different institutions. Regardless, the aim of data mining approaches is always to identify drug-symptom combinations that are interesting for further safety evaluations, for instance because they are reported more often than expected. This is an efficient but time-consuming system, since adverse reaction reports need to be digitally transcribed, usually by the national medicine centres (if not in digital format originally), coded and structured in a form that can be processed with traditional analytic tools (Lindquist 2008). Standardisation and codification are indeed an essential step to make database useful. Pharmacovigilance experts unanimously agree that "The quality of what you get from the database depends on the quality of what you put in" (Barwick 2020), an issue that the pandemic made even more visible, as we are going to describe.

During the COVID-19 pandemic, the global dimension of the therapy and vaccination programmes, together with the need for close safety monitoring of the marketed products due to scarce pre-marketing information, have generated extraordinarily large amounts of spontaneous adverse effect reports. Since January 2021, over 1.100.000 adverse effect reports of COVID-19 vaccines have been shared into VigiBase,[2] which is an unprecedented affluence. The first problem to face was that market authorisation olders and national centres are not equipped to deal with these amounts, which require more trained professionals to process the data than available at the moment. As a result, there were substantial backlogs in handling of reports even at normally resource-rich centres (Norwegian Medicines Agency 2021b).

On the other side, spontaneous reports are only part of the potentially useful data that are being produced in increasing amounts. Clinical trials, health registries, claim registries, and even experiences largely shared in social media might give insights for safety monitoring (Hussain 2021). These represent big potentials as well as big challenges. First, joining different registries, databases and health records requires expanded standardisation and a common language for coding (Leonelli 2019). Second, healthcare data are protected by privacy and cannot readily be shared among different stakeholders (Benzschawel & Silveira 2011). Third, processing unstructured data, such as clinical cases and patient narratives,

[2] Data retrieved from the website http://www.vigiaccess.org/, which provides public access to VigiBase.

requires more sophisticated analytical tools than the ones currently used to mine structured data (Hussain 2021). The issue of dealing with increased amounts of data during the pandemic, therefore, have been described predominantly as a series of practical challenges.

4.2 How to Cope with Bigger Amounts of Data? Practical Implemented Measures

The current situation has been described as an unprecedented opportunity for technological innovation (Ferreira-da-Silva et al. 2021; Hussain 2021; ICON 2020; Meng 2020; Pharmafile 2021).

Manufacturers, companies offering pharmacovigilance services, and national agencies have implemented new technologies, often based on artificial intelligence, with among them the following aims:

- Automatic coding of the adverse drug reaction into standardised medical terminology (Pharmafile 2021).
- Automatic translation into and from different languages (Pharmafile 2021).
- Increased efforts for the implementation of existing methods for automatic removal of patient sensitive data from clinical narratives, in order to share healthcare data among different databases (Meldau 2018).
- Improved mining of unstructured data, such as narratives, clinical studies and social media (Hussain 2021).

Researchers have also applied to pharmacovigilance databases data analysis methods from other disciplines, such as time series analysis (Beninger 2021). Moreover, previous efforts to link health data from different electronic registries (Hripcsak et al. 2015) have been harnessed and developed to answer COVID-19 related questions, including questions about safety of treatment (Morales et al. 2021). Finally, the European Medicine Agency, recognising that “Big Data can complement clinical trials and offers major opportunities to improve the evidence upon which we take decisions on medicines”, have set up a Big Data Taskforce to build technical skills, capacity and tools for the joint analysis of different type of data sources (European Medicines Agency 2020).

4.3 Epistemology of Big Data Pharmacovigilance: A Critical Reflection

Although the success of data-centric research is based on technological and practical innovations, it also depends on a solid base of theoretical knowledge and human judgement. Philosophical issues linked to big data are comparatively less visible in mainstream discussions but should not be overlooked. While epistemology and ethics of big data have been discussed in other disciplines dealing with big databases, such as biology and climate science (Leonelli 2016), the time is ripe for applying them to pharmacovigilance, too. The aim is to acknowledge the *full* range of skills necessary to develop an efficient use of pharmacovigilance data, in normal times and even more importantly in times of crisis.

A crucial philosophical issue to consider, when critically reflecting on the acceleration of big-data pharmacovigilance during COVID-19, is the debate between objectivity and constructivism, or else the question of theory-laden observations. The empiricist ideal that scientific explanations somehow emerge directly out of the data seems to be having a revival in era of big data (Leonelli 2016). This is in line with the evidence-based medicine paradigm, in which expertise and theory have the lowest epistemic status, and statistical evidence from controlled studies the highest (Howick 2011). Data-driven research has been saluted as 'the death of subjectivity' and is believed to lend objectivity and clarity even to fields that have been traditionally less amenable to quantification, such as sociology (McKie & Ryan 2016). Is this view, that clear explanations derive primarily from data rather than from people and expertise, applicable also to COVID-19 pharmacovigilance (and pharmacovigilance in general)?

In her analysis of data-centric biology, philosopher Sabina Leonelli writes:

> Far from being 'the end of theory', the computational mining of big data involves significant theoretical commitments. The choice and definition of keywords used to classify and retrieve data matters enormously to their subsequent interpretation. Linking diverse datasets means making decisions about the concepts through which nature is best represented and investigated. In other words, the networks of concepts associated with data in big data infrastructures should be viewed as theories: ways of seeing the biological world that guide scientific reasoning and the direction of research, which are often revised to take into account new discoveries (Leonelli 2019: 2).

We are going to show that just like theoretical understanding of natural phenomena is crucial for linking datasets in the field of big data biology, as pointed out by Leonelli, clinical and pharmacological reasoning are necessary for the meaningful organisation of data in pharmacovigilance databases. How so? And how does this matter for COVID-19 related pharmacovigilance?

We will use two examples to illustrate that the success in COVID-19 vaccine safety monitoring, although being data-driven, has not emerged directly from the data, but from a genuine collaboration between data science, pharmacological theories and clinical expertise. Our aim, in other words, is to show that big-data pharmacovigilance is theory-laden and its success in times of crisis depends on a network of different types of expertise, rather than predominantly on data science. Nurturing such network and interdisciplinary dialogue is then a central part of improving pharmacovigilance in the face of health emergency.

As a first example, consider that without proper "query" systems it is not possible to retrieve data relevant for COVID-19 specific (or any other) safety questions in an efficient way. In other words, one thing is the much-discussed technical issue of coding large amounts of data, something that seems to be possibly facilitated with artificial intelligence. Another, more fundamental need is to develop the common terminology that coders (whether human or not) use to classify the data and integrate them together (Leonelli 2019).

Let us introduce some background information before we apply them to the COVID-19 vaccine safety monitoring. When entering case safety reports in a pharmacovigilance database, marketing authorization holders and national agencies need to code the name of medicines and vaccines with a standardised international classification. One classification in use at the moment is provided in the *WHO Drug* dictionary. *WHO Drug*, created by the WHO Programme for International Drug Monitoring, is constantly updated, and the magnitude of this labour is demonstrated by the fact the big task force dedicated to maintaining it at the

Uppsala Monitoring Centre (Lindquist 2008). One *WHO Drug* feature classifies medicines based on various different and *relevant criteria*, such as their pharmacological effect, indication for treatment or metabolic pathway, in Standardised Drug Groupings (SDG) (Uppsala Monitoring Centre 2020). The SDGs are not mutually exclusive and as such any drug may be listed in several SDGs. Such grouping criteria are relevant for different purposes. For instance, a medicine manufacturer might set up a clinical trial to test a certain medicine which is metabolised by enzyme E, therefore all medicines interacting with E might interfere with the study medicine. The manufacturer then will exclude from the trial all the participants that take any of the medicines listed in the *WHO Drug* SDG of "medicines inhibiting E". This was indeed the initial purpose for setting up the SDG classification: helping *WHO Drug* users from the pharmaceutical industry to manage the inclusion-exclusion criteria in their clinical trials.

Soon enough, *WHO Drug* SDGs were repurposed and integrated in the toolbox for safety monitoring analysis (Chandler & Lagerlund 2019). Imagine for instance that I suspect that a medicine X causes a certain adverse effect because it inhibits receptor R. Being able to retrieve a group of safety reports containing medicines similar to the medicine of interest X, in that they all inhibit receptor R, is important. It allows me to check, for instance, whether there is a significant correlation with the adverse effect of interest in the total number of reports at the SDG group level. This gives an indication to support (or not) the hypothesis of mechanism.

It should be clear at this point that the ways the database can be used is determined by the types of possible 'group queries'. The more relevant the SDGs or similar groupings are for a specific purpose, the more efficient may be the data mining of coded data.

Let us now consider how SDGs were used for the safety monitoring on COVID-19 vaccines. When in need of enhanced efficiency such as during the COVID-19 pandemic, *WHO Drug* specialists created new SDGs for the new purpose of facilitating the safety monitoring of COVID-19 vaccines (Uppsala Monitoring Centre 2020). In doing so, decisions were made on how adverse effect reports related to different types of COVID-19 vaccines should be linked together. Curators made decisions about how clinical and pharmacological interactions are best "represented and investigated", in Leonelli's own words. Does it make clinical sense, for instance, that RNA-based vaccines might interact with the body in different ways than vaccines containing inactivated viruses? If so, SDGs should be grouped based on the vaccine platform. This would make it possible to easily and efficiently retrieve, for instance, all reports containing RNA-based COVID-19 vaccines together with a certain symptom and check whether there is a disproportional reporting at group level. Notice now the crucial point: the idea that the type of vaccine platform has something to say about the adverse reactions it may provoke did *not* emerge directly from the data. Rather, it is a hypothesis anchored in clinical and pharmacological thinking, obtained by reasoning about the mechanism of action of different types of vaccines and the molecular mechanisms possibly at place in the patient. The SDG example then indicates that collecting more data, and improving data technology, represent only a part of the knowledge developments necessary for COVID-19 vaccine safety monitoring.

Clinical and theoretical reasoning are fundamental for a spectrum of steps in the process of curating a pharmacovigilance database. Here is a second example. What do we mean, in statistical measures of disproportionality, that the pair vaccine-symptom is reported more than expected in the database background? Which background should be used to calculate the expected statistic? Normally, the number of reports expected if the combination happened by pure chance are calculated considering the whole database. In the case of COVID-19 vaccines, however, there might be more useful background measurements. One could choose to calculate 'background expectations' using a more relevant background, for instance using only the adverse reaction reports relative to vaccines in general. Or, to narrow it down even more, one could use as background only reports relative to vaccines for agents that access the host through airways. When narrowing down the background, one aims to detect disproportionately reported reactions that are specific for the COVID-19 vaccines, while a broader background would tend to identify reactions typical to vaccines in general. Each of these choices generate different statistical results, and there is likely no unified view on the best methodology compared to what could be considered the gold standard, the full database background. Again, the crucial point is that the reason for considering one statistical method more suitable than the others for the purpose of COVID-19 vaccine monitoring does not emerge directly from the data. If that was the case, indeed, there would be only one answer to the question of which method is the best: the answer provided by "pure facts". Rather, the method one favours to calculate disproportionality depends on clinical and pharmacological reasoning, as well as on the priorities set by different evaluating bodies.

We have argued, using examples from pharmacovigilance practice, that the data-driven approach to COVID-19 vaccine safety monitoring should be seen as constructed. Indeed, it relies on judgement, theories, and clinical/pharmacological expertise as much as on data and technological development. Why is it important to point out the fundamental role of clinical and pharmacological reasoning? The first reason, already made by Leonelli for biology data-centric research, is a question of awareness and transparency. Since the theoretical reasoning underlying data processing influence the way in which data can be used, researchers and pharmacovigilance practitioners should understand and be critical of the conceptual choices made by others, that ultimately shape their own data-based research. For instance, a recent analysis tested whether mRNA vaccines are disproportionately reported together with MedDRA terms describing facial paralysis (Kamath et al 2020). The type of statistical analysis described by the authors assumed that vaccinated and non-vaccinated people have similar likelihood of reporting an event. Evaluating whether such assumption is viable, however, is job for pharmacists and sociologists, who can assess whether for instance media campaigns might have influenced the reporting rate of vaccinated people.

A second reason for pointing out the importance of judgement and expertise in data-centric COVID-19 pharmacovigilance concerns the type of knowledge and skills we, as a scientific community, need to encourage in order to increase its efficiency, especially in times of emergency. The European Medicine Agency's Big Data Taskforce highlighted the need of more data scientists and AI professionals (European Medicines Agency 2020). However, from our arguments here stems the additional need of nurturing and reinforcing the interdisciplinary work of medical doctors, pharmacologists, and data scientists.

4.4 Ethics of Big Data Pharmacovigilance: A Critical Reflection

Increasing reliance of big data requires a parallel increase of reflections about good practices of big data research. The field of *data ethics* was recently created to study "moral problems related to data, algorithms and corresponding practices, in order to formulate and support morally good solutions" (Taddeo & Floridi 2016: abstract).

In pharmacovigilance, one dominant concern in the sphere of data ethics is the protection of patient privacy and sensitive health data (Callréus 2013). As in all epidemiological research where health data are shared between different databases, there is a tension between the potential public health benefits of accessing personal health-related information and the privacy rights of single persons (Rocca & Anjum 2020). While this tension brings about an important and still unsolved hinder to data sharing, which was also acknowledged to slow down the progression of COVID-19 data-based research (The Alan Turing Institute 2021), there is more to be discussed.

For example, we believe that some straight-forward observations about the pharmacovigilance databases should be brought to the attention of data ethicists and might raise discussions about the inclusiveness of the current system. For instance, 80% of COVID-19 related adverse reaction reports shared into VigiBase in 2020 were from the WHO regions of Europe and the Americas, and only 1% came from the African region (Rocca et al 2021). This extreme difference is more pronounced for the COVID-19 reporting than for the database and supports the observation that global differences in medicines availability and quality of healthcare have become more pronounced during the pandemic (McMahonid, Peters, Iversid & Freemanid 2020).

When considering the state of patient safety in the African continent these numbers are not surprising. Only a few countries in the WHO African region, for instance Tanzania and Ghana, have functional regulatory and pharmacovigilance systems according to international standards, and it was predicted that other governments will not be in the economic situation to prioritise pharmacovigilance in the near future (Ogar, Mathenge, Khaemba & Ndagije 2020). Arguably, the issue of limited resources is also accompanied by a language barrier. Although coding dictionaries are offered in a number of languages, pharmacovigilance protocols and reports are predominantly issued in English, something that makes it necessary for a pharmacovigilance professional to master this language. Finally, the social structures and cultural heritage of certain countries might make it less immediate for citizens to report what can be seen as a 'failure of the system', regardless of the pharmacovigilance structures in place. At the same time, regional experts warn that the COVID-19 emergency poses a particular threat to patient safety in sub-Saharan Africa, where lack of medical literacy, misinformation, lack of sufficient professional guidance in a context of panic and fear might lead to irrational use and abuse of medicines and traditional remedies to a higher extend than elsewhere, in the attempt to prevent or cure COVID-19 (Ogar et al. 2020).

It is important to highlight that when we are strengthening big-data pharmacovigilance, AI and data processing, we are representing almost exclusively European countries, the Americas and a handful of other countries globally. One side of the problem is then that global pharmacovigilance data are biased because they are incomplete. We have little information on the level of access to and the impact of COVID-19 treatments and vaccines for a large proportion of the global population.

The bias inbuilt and hidden in data-centric research is one of the dominant themes in data ethics. The concern is that cultural assumptions hold the false belief that datasets and algorithms increase objectivity of the research because they are less partial and less discriminatory than single researchers, single experiments and small datasets. Instead, it is often the case that there are inbuilt systematic discriminations, which are carried on no matter how big the datasets and how sophisticated the algorithms (D'Ignazio & Klein 2020). Although bigger studies and systematic reviews increase beliefs of objectivity due to bigger dataset, the picture is not complete until the systematic discrimination has been taken care of.

In the presented case, it seems that until social structures and inequalities are addressed, capacity building and awareness is raised and funds are allocated in order to strengthen the culture and the structures of patient safety globally, it will not be possible to at least decrease, if not overcome, the incompleteness of global pharmacovigilance data on which patient safety action is based. It seems then reasonable to argue that an increased reliance on algorithms and databases to improve drug safety needs to be accompanied by an increased effort of adapting to the social and technical structures of developing countries. Failure to do that will result in a system that contributes to increase the global inequalities of healthcare by increasing the disproportionate amounts of safety data on medicines from specific world regions.

In summary, in this session we have outlined the challenge of dealing with increased amount of data, due to the high number of drugs and vaccines with less established safety profiles that are distributed globally during the COVID-19 pandemic and potential similar future health challenges. We have shown that the pharmacovigilance community in parts of the world has implemented a number of technical innovations, based on smart algorithms and artificial intelligence, to attempt to face such challenges. Finally, we have made the point that the increased reliance on databased and algorithms must be paralleled by an increased reflection about the full manual or human skills that are necessary to make data-centric pharmacovigilance efficient in COVID-19, as well as reflections about the structural inequalities that underlie global pharmacovigilance. When such considerations are made, efforts to increase the interdisciplinarity between data-science skills and clinical expertise seem vital, together with considerations on how to improve technical know-how in developing countries.

5. Concluding Remarks

The aim of this paper was to indicate that an improvement of pharmacovigilance systems in the face of a pandemic requires the critical consideration of foundational issues at the side of technological development. Our analysis pointed out that both high uncertainty and increased focus on big data require to strengthen interdisciplinary networks between clinicians, pharmacovigilance experts, regulators, data scientists and curators of databases, data-ethicists and philosophers of science. At the moment, there is generally an increasing demand of interdisciplinary practice, however education, research funding, scientific journals and regulatory systems maintain a disciplinary focus. In particular, interdisciplinarity between the research and practice of pharmacovigilance and the humanities is still at an embryonic stage (Rocca 2020). The next question is how such interdisciplinarity should be implemented, and who is in charge of implementing it. We urge that the pharmacovigilance community should give space to this question, together with other foundational reflections on the epistemology and ethics of pharmacovigilance, in discussion fora, platforms, specialised journals and social media.

References

Ackley, T.W., McManus, D., Topal, J.E., Cicali, B. & Shah, S. 2021, "A Valid Warning or Clinical Lore: An Evaluation of Safety Outcomes of Remdesivir in Patients with Impaired Renal Function from a Multicenter Matched Cohort", *Antimicrobial Agents and Chemotherapy*, 65, 1-8.

Anjum, R.L. & Rocca, E. 2019, "From Ideal to Real Risk: Philosophy of Causation Meets Risk Analysis", *Risk Analysis*, 39, 729-40.

Barwick, M. 2020, "Pharmacovigilance in a Time of Crisis", retrieved from https:// www. youtube.com/watch?v=Agn0_FADOhM.

Beninger, P. 2021, "Influence of COVID-19 on the Pharmacovigilance Workforce of the Future", *Clinical Therapeutics*, 43, 369-71.

Benzschawel, S. & Da Silveira, M. 2011, "Protecting Patient Privacy when Sharing Medical Data", *The Third International Conference on eHealth*, *Telemedicine, and Social Medicine (c)*, 108-13.

Callréus, T. 2013, "Pharmacovigilance and Public Health Ethics", *Pharmaceutical Medicine*, 27, 157-64.

Chandler, R. & Lagerlund, O. 2019, "The Utilisation of a New Tool in Signal Management—WHO Drug Standardised Drug Groupings", retrieved from https:// www.who-umc.org/media/165059/whodrug-sdgsfinalweb.pdf

CIOMS Working Group VIII, *Practical Aspects of Signal Detection in Pharmacovigilance*, Geneva: CIOMS, 2010.

Copeland, S. 2017, "On Serendipity in Science: Discoveries at the Intersection of Chance and Wisdom", *Synthese*, 196, 2385-2406.

D'Ignazio, C. & Klein, L.F. 2020, *Data Feminism*, Cambridge, MA: The MIT Press.

Dally, A. 1998, "Thalidomide: Was the Tragedy Preventable?", *Lancet*, 351, 1197-99.

Desai, M.K. 2020, "Pharmacovigilance and Assessment of Drug Safety Reports during COVID 19", *Perspectives in Clinical Research*, 11, 128-31.

European Commission Enterprise and Industry Directorate 2009, "A Guideline on Summary of Product Characteristics", retrieved from https://ec.europa.eu/health/sites/default/files/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf.

European Medicines Agency 2020, "HMA-EMA Joint Big Data Taskforce Phase II Report: Evolving Data-Driven Regulation", 1, retrieved from www.ema.europa.eu.

Ferreira-da-Silva, R., Ribeiro-Vaz, I., Morato, M. & Polónia, J.J. 2021, "Guiding Axes for Drug Safety Management of Pharmacovigilance Centres during the COVID-19 Era", *International Journal of Clinical Pharmacy*, 43, 4, 1133-38.

Gérard, A. et alii 2021, "Remdesivir and Acute Renal Failure: a Potential Safety Signal from Disproportionality Analysis of the WHO Safety Database", *Clinical Pharmacology and Therapeutics*, 109, 1021-24.

Gevers, S., Welink, J. & van Nieuwkoop, C. 2021, "Remdesivir in COVID-19 Patients with Impaired Renal Function", *Journal of the American Society of Nephrology*, 32, 518-19.

Grandvuillemin, A., Drici, M.D., Jonville-Bera, A.P., Micallef, J. & Montastruc, J.L. 2021, "French Pharmacovigilance Public System and COVID-19 Pandemic", *Drug Safety*, 44, 405-408.

Howick, J. 2011, *The Philosophy of Evidence-based Medicine*, Oxford: Wiley-Blackwell, BMJ Books.

Hripcsak et alii 2015, "Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers", *Studies in Health Technology and Informatics*, 216, 574-78.

Hussain, R. 2021, "Big Data, Medicines Safety and Pharmacovigilance", *Journal of Pharmaceutical Policy and Practice*, 14, 48.

Kamath, A., Maity, N. & Nayak, M.A. 2020, "Facial Paralysis Following Influenza Vaccination: A Disproportionality Analysis Using the Vaccine Adverse Event Reporting System Database", *Clinical Drug Investigations*, 40, 883-89.

ICON. 2020, "The Impact of COVID-19 on Pharmacovigilance", retrieved from https://www.iconplc.com/insights/blog/2020/04/29/the-impact-of-covid-19-on-pharmacovigilance/.

Leonelli, S. 2016, *Data-Centric Biology. A Philosophical Study*, Chicago and London: The University of Chicago Press.

Leonelli, S. 2019, "The Challenges of Big Data Biology", *ELife*, 8, 1-5, doi: 10.7554/eLife.47381.

Leonelli, S. 2021, "Data Science in Times of Pan(dem)ic", *Harvard Data Science Review*, 3, 1-30.

Letourneau, M., Wells, G., Walop, W., & Duclos, P. 2008, "Improving Global Monitoring of Vaccine Safety: A Survey of National Centres Participating in the WHO Program for International Drug Monitoring", *Drug Safety*, 31, 389-98.

Lindquist, M. 2008, "VigiBase, the WHO Global ICSR Database System: Basic Facts", *Drug Information Journal*, 42, 409-19.

Machhi, J. et alii 2020, "The Natural History, Pathobiology, and Clinical Manifestations of SARS-CoV-2 Infections", *Journal of Neuroimmune Pharmacology*, 15, 359-86.

McKie, L. & Ryan, L. 2016, *An End to the Crisis of Empirical Sociology? Trends and Challenges in Social Research*, London and New York: Routledge.

McMahonid, D.E., Peters, G.A., Iversid, L.C. & Freemanid, E.E. 2020, "Global Resource Shortages during Covid-19: Bad News for Low-income Countries", *PLoS Neglected Tropical Diseases*, 14, 1-3.

Meldau, E.L. 2018, "Deep Neural Networks for Inverse De-Identification of Medical Case Narratives in Reports of Suspected Adverse Drug Reactions. Degree Project Computer Science and Engineering", retrieved from http://www.diva-portal.org/smash/get/diva2:1185934/FULLTEXT01.pdf

Meng, X. 2020, "COVID-19: A Massive Stress-test with Many Unexpected Opportunities (for Data Science)", *Harvard Data Science Review*, Special issue 1-COVID-19, retrieved from https://hdsr.mitpress.mit.edu/pub/l7a2t45s/release/1?readingCollection=0181d53b

Meyboom, R.H., Hekster, Y.A., Egberts, A.C, Gribnau, F.W. & Edwards, I.R. 1997, "Causal or Casual? The Role of Causality Assessment in Pharmacovigilance", *Drug Safety*, 17, 6, 374-89.

Michener, W.K. et alii 2009, "Biological Field Stations: Research Legacies and Sites for Serendipity", *BioScience*, 59, 300-10.

Morales, D.R. et alii. 2021, "Renin-Angiotensin System Blockers and Susceptibility to COVID-19: An International, Open Science, Cohort Analysis", *The Lancet Digital Health*, 3, 2, doi: 10.1016/S2589-7500(20)30289-2.

Mugosa, S., Stankovic, M., Turkovic, N. & Sahman-Zaimovic, M. 2015, "Medical Dictionary MedDRA: Used in over 60 Countries, among which is Montenegro", *Hospital Pharmacology - International Multidisciplinary Journal*, 2, 266-71.

Mühlberg, W. & Platt, D. 1999, "Age-Dependent Changes of the Kidneys: Pharmacological Implications", *Gerontology*, 45, 243-53.

Naidu, M.V.S., Sushma, D.S., Jaiswal, V. & Asha, S.P.T. 2020, "The Role of Advanced Technologies Supplemented with Traditional Methods in Pharmacovigilance Sciences", *Recent Patent on Biotechnology*, 14, 1, doi: 10.2174/1872208314666 201021162704.

Norwegian Medicines Agency 2021, "Norwegian Medicines Agency Notified of Blood Clots and Bleeding in Younger People after Vaccination with AstraZeneca Vaccine", retrieved from https://legemiddelverket.no/nyheter/norwegian-medicines-agency-notified-of-blood-clots-and-bleeding-in-younger-people-after-vaccination-with-astrazeneca-vaccine.

Norwegian Medicines Agency 2021b, "Reports of Possible Adverse Reactions to COVID-19 Vaccines as of 31 August 2021", retrieved from https://legemiddelverket.no/nyheter/reports-of-possible-adverse-reactions-to-covid-19-vaccines-as-of -31-august-2021.

Ogar, C., Mathenge, W., Khaemba, C. & Ndagije, H. 2020, "The Challenging Times and Opportunities for Pharmacovigilance in Africa during the COVID-19 Pandemic", *Drugs and Therapy Perspectives*, 36, 351-54.

Osimani, B. 2013, "Until RCT Proven? On the Asymmetry of Evidence Requirements for Risk Assessment", *Journal of Evaluation in Clinical Practice*, 19, 454-62.

Pharmafile 2021, "How COVID-19 Has Changed Pharmacovigilance", retrieved from http://www.pharmafile.com/news/571507/how-covid-19-has-changed-pharmaco vigilance.

Rocca, E. & Andersen, F. 2017, "How Biological Background Assumptions Influence Scientific Risk Evaluation of Stacked Genetically Modified Plants: An Analysis of Research Hypotheses and Argumentations", *Life Sciences, Society and Policy*, 13, doi: 10.1186/s40504-017-0057-7.

Rocca, E., Anjum, R.L. & Mumford, S. 2017, "Post-Marketing Risk Assessment of Drugs as a Way to Uncover Causal Mechanisms", in La Caze, A. & Osimani, B. (eds.), *Uncertainty in Pharmacology: Epistemology, Methods and Decisions*, Boston Series for the History and Philosophy of Science, Cham: Springer.

Rocca, E., Copeland, S. & Edwards, I.R. 2019, "Pharmacovigilance as Scientific Discovery: An Argument for Trans-Disciplinarity", *Drug Safety*, 42,1115-24.

Rocca, E. & Anjum, R.L. 2020, "Erice Call for Change: Utilising Patient Experiences to Enhance the Quality and Safety of Healthcare", *Drug Safety*, 43, 513-15.

Rocca, E. 2020, "Philosophy of Science Meets Patient Safety", *Uppsala Reports*, 82, 16-19.

Rocca, E. et alii 2021, "Remdesivir in the COVID-19 Pandemic: An Analysis of Spontaneous Reports in Vigibase during 2020", *Drug Safety*, 44, 987-98.

Saint-Raymond, A. et alii 2020, "Remdesivir Emergency Approvals: A Comparison of the U.S., Japanese, and EU Systems", *Expert Review of Clinical Pharmacology*, 13, 1095-1101.

Shakir, S.A.W. & Layton, D. 2002, "Causal Association in Pharmacovigilance and Pharmacoepidemiology", *Drug Safety*, 25, 467-71.

Taddeo, M., & Floridi, L. 2016, “What Is Data Ethics?”, *Philosophical Transactions of the Royal Society A*, 374, 1-5.

The Alan Turing Institute 2021, “Data Science and AI in the Age of COVID-19”, retrieved from https://www.turing.ac.uk/sites/default/files/2021-06/data-science-and-ai-in-the-age-of-covid_full-report_2.pdf.

Trontell, A. 2004 “Expecting the Unexpected—Drug Safety, Pharmacovigilance, and the Prepared Mind”, *New England Journal of Medicine*, 351, 1385-87.

Uppsala Monitoring Centre 2020, “WHO Drug Newsletter December 2020”, retrieved from https://www.anpdm.com/newsletterweb/464A5D4A77454A58447 4494659/43445E437440415143794141 5D4171.

Uppsala Monitoring Centre 2021a, “What is a Signal?”, retrieved from https://www.who-umc.org/research-scientific-development/signal-detection/what-is-a-signal/.

Uppsala Monitoring Centre 2021b, “The Use of the WHO-UMC System for Standardized Case Causality Assessment”, retrieved from http://www.who-umc.org/graphics/4409.pdf.

Models and Experts: The Contribution of Expertise to Epidemic and Pandemic Modelling

Carlo Martini

Università Vita-Salute San Raffaele

Abstract

Modelling is a precious source of information in science. With models, we can simplify an otherwise messy reality in order to understand the fundamental driving forces of a system, like an epidemic, and we can try to predict the course of events in complex scenarios where there is a great degree of uncertainty. In short, models can be used to explain and predict phenomena. Yet models interact with expert opinions in two fundamental ways. They are sometimes in competition with expert opinion, and they are sometimes heavily dependent, for their proper working, on expert opinion.

In this paper I will illustrate the different ways in which a model interacts with expert opinion. I will focus on epidemiological models. I will explain how, in epidemic modelling, getting the expertise right is as important as getting the model right. I will briefly present epidemiological models with a focus on the specific contribution of expert judgment to the choice and use of these models. I will compare expert judgment with statistical judgment, highlighting the limits of the former. I will analyse the interconnectedness of modelling and expert judgment in epidemic simulations based on a case report and, finally, I will suggest some strategies for ameliorating the interaction between modelling and expert judgment.

Keywords: Epidemiological modeling, Expertise, Expert judgment, Statistical judgment.

1. Introduction

Modelling is a precious source of information in science. With models, we can simplify an otherwise messy reality in order to understand the fundamental driving forces of a system, like an epidemic. How different would the spread of a virus look like, if the driving mode of transmission was through airborne particles, or droplets, or fomites? With models, we can try to predict the course of events in complex scenarios where there is a great degree of uncertainty. Mill compared studying social phenomena to studying the laws of tides; we can only aim for approximation and inexactness when the course of a natural or social phenomenon is determined not

ISSN 2465-2334
DOI 10.14275/2465-2334/202113.mar

by a few generally well-known factors, but by a complex interaction of many causal factors:

> circumstances of a local or casual nature, such as the configuration of the bottom of the ocean, the degree of confinement from shores, the direction of the wind, &c., influence in many or in all places the height and time of the tide; and a portion of these circumstances being either not accurately knowable, not precisely measurable, or not capable of being certainly foreseen, the tide in known places commonly varies from the calculated result of general principles by some difference that we cannot explain, and in unknown ones may vary from it by a difference that we are not able to foresee or conjecture (Mill 1882: 587).

I will come back to this passage of Mill's work in due time, because it tells us something about the difficulties of modelling an epidemic. In short, models can be used to explain and predict phenomena. Allegedly, there are other purposes too, but for now I will stick to these two. Models work by isolation and idealization (Mäki 1992). For example, epidemiological models can give us at best a general idea of how a virus could spread if we make several simplifying assumptions. Most importantly, models often fail in performing their explanatory and predictive functions without expert judgment. The assumptions that go into a model, its parametrization, and its connection with a target system are all dependent in multiple ways on expert judgment.

In the rest of this paper, I will highlight the different phases in which expert judgment affects the development and application of a model. I will focus on epidemiological models.[1] While explanation is an important component of these types of models, epidemiological models are important because they are used to predict the spread of viruses in epidemic or pandemic situations given a range of pharmaceutical (e.g., antiviral drugs) and non-pharmaceutical (e.g., social distancing) policy measures that a society will usually put in place to respond to a pandemic or epidemic scenario. I will explain how, in epidemic modelling, getting the expertise right is as important as getting the model right. In the next section I will briefly present epidemiological models with a focus on the specific contribution of expert judgment to the choice and use of these models. In section 3 I will compare expert judgment with statistical judgment, highlighting the limits of the former. Section 4 will analyse the interconnectedness of modelling and expert judgment in epidemic simulations based on a report by the *National Academies of Sciences, Engineering, and Medicine* published (Institute of Medicine 2006). Section 5 will suggest some strategies for ameliorating the interaction between modelling and expert judgment and section 6 will conclude.

2. Expert Input in Epidemic Modelling: A Primer

Epidemic and pandemic models are usually systems of equations, often implemented in a computer programme. In Black's typology, they are mathematical

[1] In this paper I will use the term to specifically indicate mathematical modelling of infectious diseases—that is, that class of epidemic models that describe the spread of an infectious disease.

models (Black 1962), and they are theoretical models according to Achinstein's typology (1968):[2]

> The use of such a model characteristically involves the awareness and explicit acknowledgement that the real object is far more complex than its representation in the model: the theoretical model assumes away many complications while highlighting limited aspects of the object (Mäki 2001: 9932).

Indeed, epidemics and pandemics are complex phenomena: the spread of a virus depends on a very large number of factors, including biological aspects of the virus itself, of the range of hosts it can infect, and, as far as human health is concerned, it depends on human behaviour and available technology (e.g., availability of antiviral therapies and vaccines).

In this paper I focus on epidemiological models of two principal kinds: SIR models, and SIS models. The acronyms stand, respectively, for *Susceptible, Infectives, Removed*, and *Susceptible, Infectives, Susceptible*, the main difference being that in the latter model a virus can reinfect a host who has previously been infected. Depending on whether infection confers immunity in the hosts that survive the infection, the SIR and SIS models divide the total population in two or three macro subpopulations. At any given point in time a population will have three types of hosts: the susceptible are those who can be infected by the virus, the infective are those who are infected with the virus, and the removed are those who are either immune or dead, after infection.

Most viruses have a certain rate of reinfection, so model choice depends on knowledge about the virus and how it interacts with the host's immune system. Knowledge acquired in the early phases of a pandemic is very precious because it helps determine what kind of model we should be using, even before we think of the next steps, like parametrization and goodness of fit. For instance, in the early phases of the COVID-19 pandemic, sporadic cases of reinfection caused uncertainty about whether immunity from SARS-CoV-2 lasted more than a few weeks after recovery. As research progressed, scientists were able to determine relatively reliable rates of reinfection for different age groups, thus making model-choice easier.

Uncertainties about reinfection, and how it affects model choice, is just one example of how much modelling a pandemic needs reliable methods for collecting data. In a perfect world, we would be able to collect the information we need for model choice and parameterization in two ways: either (a) with instruments, for example, in the same way in which we measure the temperature, or pressure, in a patient, or (b) by widespread scientific consensus; for example, when we need to know the temperature of the sun we can rely on a reasonable level of scientific agreement and error rates. The world of pandemics is not perfect, and, instead, the input source for much of the knowledge that is needed is expert judgment, that is, the informed guesswork of specialists in certain areas of science. Accord-

[2] I have used mostly the term "epidemic modelling" in this paper, even though I may sometimes use the term "pandemic modelling" instead. The main difference is that epidemiological models usually contain primarily epidemiological elements, while a pandemic model may include social, economic, and other variables. The usage is not always consistent in the literature and for the scope of this paper any differences between the two can be ignored.

ing to Cooke (1991: 30) "musings, brainstorms, guesses, and speculations of experts can be significant input in a structured decision process". In epidemiological modelling, musing, guesses, etc. are often the main source of modelling input we have.

There are two possible interactions between modelling and expert judgment: Expert judgment can be a constitutive part of modelling—for example when using expert judgment as input for the parameters of the model—or expert judgment can be an alternative to model-based judgment—for example when the model is inadequate to represent the problem at hand. The following sections will deal with both of these scenarios in which there is interaction, and sometimes a conflict, between a model and an expert judgment.

Next, I list the main contribution of expert judgment to pandemic and epidemic modelling.

A. **Model Choice**: I have already illustrated this point above. It's important to note that the choice of the modelling framework is heavily dependent on epidemiological knowledge of the interaction between the organism and the infected host. For example, reinfection possibility (and rates) can significantly change the dynamics of the model. Other sources of model uncertainty are the choice of variables, the degree of detail, and the endemic dynamics that the model is designed to capture. Much of the knowledge needed to reduce model uncertainty can only be obtained by consulting experts, even though reinfection rates can be derived statistically, provided we have enough data. Whether this information can be obtained by reliable methods depends on the type of infectious disease we are considering. The older and the more data we have, the more likely it will be that a scientific consensus has formed around key assumptions. With pandemics as recent as COVID-19, especially in the early waves, relevant knowledge comes from the informed speculations of experts, and by comparison with data from similar viruses (e.g., other coronaviruses) and pandemics (e.g., SARS, MERS). Analogical reasoning is highly fallible and dependent on human reasoning, that is, expert judgment: "It is a fact about human cognition that we very commonly make a judgment that one case is similar to another in drawing conclusions about what to do in daily life" (Walton et al. 2008: 55).

B. **Parameter Selection and Parametrization**: One of the fundamental sources of uncertainty in models is parameter uncertainty. The list of parameters that are theoretically relevant to an epidemiological model is virtually endless. Even though parsimonious models containing only a few variables are often considered to be more valuable to isolate key features of a pandemic (Bertozzi et al. 2020), we still need experts to make a judgment of relevance in the first place. For example, the connectedness of a network structure is fundamental for understanding how many steps a virus needs to spread through an entire network, given the same number of nodes (hosts). Social habits, geographical distribution of hosts, and many other factors affect a network's connectedness. The same is true about the clustering of a network (in lay terms, the layout of a network and how its nodes are distributed), and its degree of centrality (whether there are nodes in the network that are connected to all or most of the other nodes). Other important parameters are, for example, the rate of infectiousness—the R-number that for COVID-19 made headlines time and again (Adam 2020)—and the mode of transmission, namely through aerosol, droplets or fomites. Parameterization of a model often needs significant expert input. Some of the parameters can be obtained by statistical calculation; for instance, the Basic Reproduction Number can be relatively straightforwardly obtained via a variety of estimation methods. Other parameters, however, are harder to estimate, like compliance with behavioural interventions. Other the parameters can only be estimated with significant uncertainty: Morse et al. (2006) highlight the significant amount of uncertainty for very important parameters (e.g., effectiveness of non-pharmacological practices) and state that there are no science-based compendia of best practices.

3. Expert Judgment and Statistical Judgment

At this point I must clarify the difference between expert judgment and statistical (also mechanical, or actuarial) judgment. Expert judgment refers to the judgment of a human expert. I cannot provide an account of expertise in this article, and I will generally refer to experts as those who have attained sufficient experience and competence in a relatively narrow field of human knowledge (Martini 2019). Experts are said to possess tacit knowledge, through the application of which they are able to perform tasks (know-how), or give answers to well-formulated problems (know-what). For example, expert forecasters can answer questions like "what is the probability of a 48-hour clear-weather window for the next seven days on Everest".[3] For obvious reasons, in this paper I am considering only know-what experts.

Statistical judgment, on the other hand, is judgment delivered by calculation. We can think of expert systems as an example of statistical judgment. An expert system is a system of rules that can be implemented into a computer to aid or substitute human decision-making (Dreyfus 1987). Medical diagnoses are examples of decision problems: what is the most likely diagnosis for patient X, given symptoms Y_1, Y_2, …, Y_n, and patient characteristics Z_1, Z_2, …, Z_n? To mention a few concrete cases, Seixas et al. (2014) develop an expert system to support the diagnosis of a range of neurological disorders, Samuel and Omisore (2013) give a Web-Based Decision Support System for typhoid fever, and the list could become very long. Medical expert systems are fed data, either automatically (e.g., patient temperature) or through an operator (is the patient experiencing shivers?) and use algorithms to reach a conclusion on a diagnostic query.

Since the 1950s, Paul Meehl and, later, his collaborators, have undertaken an extensive research programme to show the superiority of statistical judgment over expert judgment. In the field of psychiatric diagnostics, Meehl showed that simple algorithmic tools were often able to outperform clinical evaluation in predicting human behaviour. A simple example will illustrate: Let us imagine that our task is to distinguish between psychotic and neurotic patients. We have two options: a) a clinical evaluation where a physician examines the patient; b) the Goldberg Rule. In 1965 Goldberg's study showed that a simple actuarial rule was

[3] The example is not random: The world of alpinism regards Karl Gabel and Vitor Baía as two of the best weather forecasters for high-altitude mountaineering. They have provided forecasts to top-class alpinists during their expeditions around the world. Their forecasts rely on models and data but also on significant experience and knowledge about mountaineering and the behaviour of weather patterns around high mountains. See Benavides 2018.

performing better than clinicians in diagnosing patients as either psychotic or neurotic. The rule takes a number of inputs from validity and clinical scales and gives a diagnosis based on the outcome of a simple calculation.

$$X = (L + Pa + Sc) - (Hy + Pt)$$

[L is a validity scale and Pa, Sc, Hy, and Pt are clinical scales of the MMPI: in order Paranoia, Schizophrenia, Hypomania, Psychasthenia]

If $x < 45$, diagnose patient as neurotic. If $x > 45$, diagnose patient as psychotic.

Goldberg Rule (from Bishop and Trout 2005: 14)

Goldberg's rule doesn't need significant judgment input. A short training, or even a user-friendly interface, will be enough for asking the patient the right questions and collecting the right data, and the rest is left to the algorithm. According to Bishop and Trout, "Sometimes, it would be better for the experts to hand their caseload over to a simple formula that a smart 8-year-old could solve" (Bishop and Trout 2005: 25).

By looking at much of the literature on the benefits of statistical over expert judgment, we may be tempted to think that, for most or possibly all decision problems, computers and algorithms perform better than human judgment. The claim is probably true, but it comes with a very important caveat: it is true for most phenomena for which we either have a rather precise understanding of the mechanisms involved, or enough statistical data. Unfortunately, not all phenomena are of this kind.

Aspinall and Cooke provide a fitting example of a decision problem for which expert judgment is needed; namely, predicting volcano behaviour:

> When a potentially deadly volcano becomes restless, civil authorities invariably turn to scientific specialists to seek to anticipate what the volcano will do next, and to help them judge the danger. Although it is usually possible to discern the earliest signs of unrest, *forecasting the course and precise timing of eruptions* still remains awkwardly inexact (Aspinall and Cooke 1998: 2113, italics added).

Aspinall and Cooke's article explains the application of the Cooke Method of expert elicitation to the Soufrière Hills volcano, a previously dormant volcano that became active in 1995 in the populated island of Montserrat and has been claiming land and lives ever since. They state that, despite the extensive monitoring equipment the Montserrat Volcano Observatory has set up all around the active volcano, expert judgment is still the predominant source of actionable information on evacuation decisions. "Even though armed with arrays of sophisticated monitoring equipment, the scientists working on the problem have a wide range of opinions about what the volcano might or might not do next" (Aspinall and Cooke 1998: 2114). A major problem with using expert judgment, however, is that it tends to produce a lot of noise. Experts tend to disagree a lot, and, while harnessing the power of their tacit knowledge, they also suffer a wide range of biases that affect human judgment (Faust 1984). In the Montserrat case, the attempt was to use Cooke's methodology of expert judgment elicitation and aggregation (Cooke 1991) to reduce the noise from experts.

In the case of the Montserrat Volcano, expert judgment filters and fills in the knowledge gaps from models about volcano behaviour and data collected on the ground. Is it possible that one day we might be able to develop a rule for deciding whether to evacuate an area on the basis of a number of suitable inputs and an underlying algorithmic rule that can process the data? Most likely so, however, as of today, we need judgment based on expert's experience and tacit knowledge.

To conclude this section, I shall return to the topic of epidemic modelling, after the digression on expert judgment and statistical judgment. Does predicting the course of an epidemic, and the effectiveness of different containment measures, look more like predicting the behaviour of a volcano, or like diagnosing psychotic and neurotic patients? In the next sections, we will try to understand how much of epidemic modelling is dependent on hard data, and how much is dependent on expert judgment. I will illustrate the contribution of expert judgment in epidemiological modelling focusing on a 2006 report of the *National Academies of Sciences, Engineering, and Medicine*.

4. Using Expert Judgment to Forecast the Next Pandemic

The 2006 report *Modeling Community Containment for Pandemic Influenza* (Institute of Medicine 2006) illustrates very well the interaction between statistical and expert judgment in epidemiological modelling. One of the report's tasks was to assess "the quality of existing models about a potential influenza pandemic and their utility for predicting the effects of various community containment policies on disease mitigation" (Institute of Medicine 2006: 1). The motivation for the report was that in 2006 experts were aware that a major pandemic was to be expected, and the report is particularly interested in evaluating the ability of models to predict the effects of nonpharmaceutical interventions (NPIs, like social distancing and face masks) in mitigating future expected outbreaks. The report stresses that the models "should be viewed as aids to decision-making, rather than substitutes for decision-making" (Institute of Medicine 2006: 4).

In modelling an epidemic there is much model uncertainty to begin with. Model uncertainty refers to the intrinsic uncertainty about the architecture of the model we choose, and the report indicates that epidemiological models need to rely "heavily upon expert judgment as to the inherent reasonableness of the model as a representation of reality" (Institute of Medicine 2006: 4). In general, reducing parameter uncertainty is less dependent on expert judgment when it is possible to run robustness analyses or collect more data. Even then, however, robustness analysis and data-gathering have costs in terms of resources and time. Good expert judgment can then reduce the need of collecting additional data and testing for robustness.

Conscious modelers are well-aware of the dependency of their data-crunching machines on expert input, and of the limits that that implies for the quality of knowledge obtained through models. One way to reduce the input from experts is to set up automatic data feeding mechanisms:

> One way to improve predictive ability is to adapt or construct decision-aid models that can incorporate surveillance data in real time and adapt to the actual experiences of an outbreak as it occurs. *Current models are based on educated guesses for a range of plausible values based on information from previous pandemics*. As a result, they are not able to predict with any certainty the future course of a pandemic and the

> effectiveness of interventions to reduce transmission (Institute of Medicine 2006: 13, italics added).

One way of reducing the dependency of models on expert input and educated guesses is then to set up mechanism for feeding surveillance data directly to the model. Nonetheless, this set up has limitations.

As in the case of volcano behaviour, it is possible and desirable that in the future much of the data the models need could be obtained through automated or semi-automated means. "Automated" means that the model can access large databases maintained by governments or private institutions, in this context it is important that data be transparent and, as much as possible, open (Molloy 2011). "Semi-automated" means that the model cannot directly access the data, but that data is nonetheless straightforwardly available to the modelers in the same way as national GDP, employment, or mortality rates are readily available. The objectivity of data that can be tapped into by a model does not need to be uncontroversial, there can still be some uncertainty; but the fundamentals of the methodology with which experts arrive at the data are shared by the large majority of the established experts.

Even once we establish a methodology for gathering and collecting better data, as Recommendation 7 of the Institute of Medicine (2006) report suggests, we are left with the problem that modelling a pandemic is an iterative process. The course of a pandemic is highly dependent on geography, human behaviour, and of course the evolution of diseases (e.g., due to virus mutation). It is unlikely that there will be, in the short run, a highly accurate predictive model like a descriptive model of celestial mechanics, because different equilibria between infective-agents and infected-hosts are likely to arise in response to changes in all the macro areas listed above: the environment, social behaviour, and the evolution of biological organisms. For that and other reasons models need to be regularly updated with new knowledge from a range of disciplines: biological and social ones. Recommendation 6 of the report reads as follows:

> The committee recommends that policymakers regularly convene forums for public dialogue on pandemic influenza modeling and analyses, and recommends the development of a *standing expert panel* to provide ongoing advice regarding models of pandemic influenza (Institute of Medicine 2006: 13).

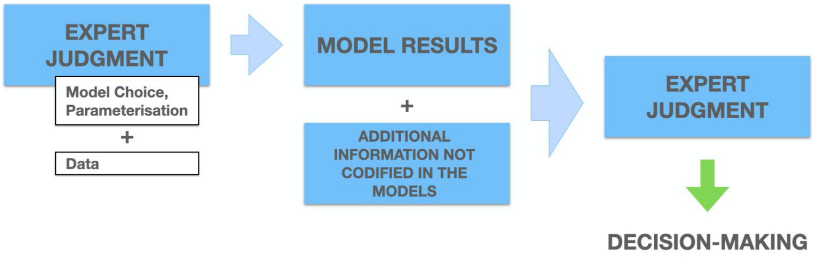
Ultimately, current epidemiological modelling, and especially the modelling of pandemic situations, relies heavily on the use of expert judgment at various phases of the modelling process. The last point I will discuss in this section, in relation to expert judgment, is the comparison between model-based evidence and expert-based evidence. The idea of expert-based evidence is that we can use experts and their judgments as estimators. Instead of relying on algorithm-driven estimates, we can use subjective probabilistic judgment (Cooke 1991). Once subjective judgments are elicited, possibly in a systematic way, they can be compared to model-based results for independent validation. This was the idea behind the work by Bankes, Aledort and colleagues, from the RAND corporation (see Institute of Medicine 2006: 9, RAND model). Aledort and colleagues ran an elicitation exercise to evaluate a package of non-pharmaceutical interventions, among which respiratory etiquette, hand hygiene, N95 respirators, etc. (see Aledort et al. 2007). The goal was to evaluate a package of non-pharmaceutical interventions during a seasonal influenza epidemic. The same package of choices was then run with two models—namely epidemiological and policy effectiveness models—and the package of preferred expert choices was compared to the results from 1000 runs of the linked models (Institute of Medicine 2006: 9).

The reason for resorting to subjective expert opinion was clear:

> In light of the evident lack of scientific evidence about specific non-pharmaceutical interventions in the context of seasonal or pandemic influenza, there was limited directly useable information from the majority of the studies identified in the formal Medline search. For this reason, *we turned to expert opinion to inform and categorize the findings* (Aledort et al. 2007: 3, italics added).

In this section I have explained how subjective expert judgment and modelling exercises for forecasting epidemiological phenomena and related containment measures are interlinked with one another. Given the complexity of the problem, the relative lack of standardized and agreed upon modelling framework, the lack of standardized procedures for collecting data, and the developing nature of the interaction between infectious pathogens and hosts, the starting point of understanding a specific pandemic situation is expert judgment. The role of epidemiological models is therefore twofold: one the one hand, simple models can help isolate and highlight key features of a phenomenon. In this respect, a model cuts out a piece of reality and allows modelers to analyse some of the chosen factors in isolation from the "messiness" of the real world. On the other hand, a computational model serves as a computational aid to otherwise intractable systems of equations, thereby making possible simulations of likely scenarios and outcomes, given varying initial conditions.

Ultimately, the results of a modelling exercise are dependent on the accuracy of the initial assumptions (model choice, and parameterization), and on the degree of uncertainty of each choice of variables and associated values. This type of complex post-hoc assessment is again in the hands of subjective expert judgment. Policy-making based on modelling results is neither straightforward nor independent of human judgment. In short, we can confidently claim that the starting and end points of modelling a pandemic are subjective expert judgments. The figure below illustrates this point.


EXPERT JUDGMENT
Model Choice, Parameterisation
+
Data
MODEL RESULTS
+
ADDITIONAL INFORMATION NOT CODIFIED IN THE MODELS
EXPERT JUDGMENT
DECISION-MAKING

The report on pandemic modelling reflects well Mill's point on studying complex phenomena like tides. We can see Mill's point in relation to a contrast class: the study of the movement of celestial bodies: astronomy. In astronomy, we can isolate a few causal factors, of similar nature to one another, and all acting in ways that can be relatively easily aggregated and taken into account. A small variation in parametrization is likely to make a difference if our instruments are good enough to detect the difference. Uncertainty and inexactness in astronomy is a matter of degree and precision of our instrumentation. Not so in the study of tides and epidemiology: tidal levels and the spread of diseases are dependent on causally interconnected factors that are very different in nature from one another, and small and large uncertainties for each of these factors can produce significantly different observations.

5. How to Handle Expert Judgment: Strategies

So far, I have assumed that the interaction between model-based and expert-based judgment is unproblematic. In this section I will explain why it is not. The main problem with expert judgment is that it tends to be biased in several ways. This is not to say that statistical judgment cannot be biased as well, but in this section I will focus on the specific ways in which biases in expert judgment reduce the epistemic quality of knowledge dependent on it. I will suggest established methodologies for reducing that bias, in particular: 1) the progressive substitution of subjective expert judgment with appropriate statistical methods, when possible, and 2) the reduction of subjective bias by means of aggregation techniques and structured expert judgment elicitation methods.

Since the work that Paul Meehl undertook in the 1950s, and the subsequent research that his *Disturbing Little Book* spurred (Meehl 1986), many scholars have highlighted the fact that human judgment is particularly ill suited at handling complex information and statistical data. Meehl, and later his collaborators, showed that those we call experts are often not so good at giving us good judgment on a number of practical questions. Trout and Bishop (2005) give the example of parole boards judging whether criminals should be eligible for parole. The heuristics and biases programme (Tversky and Kahneman 1974) has shown how human judgment fails to produce good estimates in a wide range of common tasks. The heuristics that we use to solve common daily problems, while useful with relatively menial tasks, can lead us into traps when the number of variables increases, and probabilistic interactions substitute deterministic causal pathways. As I explained above, pandemics and, in general, epidemiological phenomena, have all the characteristics of complex phenomena. The spread and impact of a virus on a host population depends on the interrelation of biological, ecological, and behavioural factors. Reliance on expert judgment, then, while it is inevitable because of our current state of knowledge, is also problematic.

The first important point here is that I have so far assumed that interrelation of model-based and expert-based judgment is straightforward. It is not. Significant literature on the interrelation has argued that whenever statistical and actuarial judgment is available, expert judgment will tend to ignore important insights from it and diminish the epistemic value of a combined judgment. In other words, when given the chance, the expert will tend to favour their own judgment, rather than the model's judgment, and often this will lead to an inferior overall assessment (Leli and Filskov 1984).

Trout and Bishop call the strategy of deviating from statistical judgment, when given the possibility, *epistemic exceptionalism* (Trout and Bishop 2005: 43). The reasoning goes as follows: let us suppose we have an algorithmic decision-making rule that works under most conditions. Under condition *X*, the rule tells us to choose *Y*. Clearly, there will be circumstances (sometimes called *broken leg* *scenarios*) under which the rule gives us the wrong answer, and we should deviate from its results. How do we know whether the rule is working or not, in a particular case? The judgment of an expert will have to provide that kind of information and, not surprisingly, experts tend to misjudge how often a case is an example of the class of cases in which the rule does not apply. Experts, that is, overestimate the number of cases in which a rule fails to provide the correct answer. Epistemic exceptionalism, therefore, is often the wrong strategy, when mixing actuarial and subjective decision-making. Bishop and Trout are clear on this: when a statistical method (an expert system) is available, experts should never stray from it except under very exceptional circumstances.

Suggestion 1. The first strategy for ameliorating our modelling practices is to improve our models in order to avoid excessive interactions between actuarial and subjective decision making. We should identify as clearly as possible those subdomains in which model-based judgment can be shown to be conclusively outperforming expert estimations. That will leave out those parts of the problems where, instead, we must rely on subjective judgment. Avoiding epistemic exceptionalism in key aspects of pandemic modelling should help clarify when subjective expert judgment ought to leave room to statistical judgment, a sort of "expert humility". The important point here is to understand that models are better at doing some things, and when we can identify the domains in which they are better we should let them handle the work.

The previous suggestion leaves implicit something I stated in the previous sections: there are domains in which models are either worse at producing knowledge, or not available at all. In sections 3 and 4 I argued we cannot predict the course of a pandemic with and without containment measures with models alone, so we must resort to established expertise. The question then is what we can do to reduce bias. For example, in the case of the COVID-19 pandemic, from the very beginning the main type of expertise that policy makers and the public listened to was health-centred. In the urgency of the initial steps that made sense because hospitals were being overrun and knowledge about the virus and possible treatments was scarce. Much thinking that went into the formulation of policies was also health-centred. This is not a note of criticism, but rather something to take at face value: Doctors, epidemiologists and health officials were in the public spotlight; their words were being analysed, criticised or otherwise glorified in the media. In that context health-centred thinking rightly influenced initial containment policies: "Up until late April, the Finnish government followed a script written predominantly by THL. THL is, by definition, a health utilitarian agency" (Häyry 2021: 117) The same is true for most governments around the world. In the medium and long run, it is possible to argue that health-centeredness can act as an epistemic bias and lead to groupthink. Groupthink is recognized as a cognitive bias that affects the quality of decision-making (Cleary et al. 2019). It is reasonable to argue that in the long run diversity of relevant expertise is important in modelling an epidemic.

Groupthink is an example of the possible biases that affect expert judgment, also in interaction with modelling efforts. The question is then whether and how expert judgment can be debiased. There is extensive literature on the subject, so in these final sections I will only be able to mention a few general points.

Suggestions 2. The second strategy for ameliorating our modelling practices is to reduce bias in expertise. There are a few ways to reduce biases:

A) **Prefer groups rather than individuals**. Committees have been shown to avoid some of the biases that would otherwise affect judgment, for example, when it comes to estimation, groups of experts tend to outperform individual experts. Cooke (1991) and his continuous work with various other collaborators, has developed a methodology of expert elicitation based on teams and judgment aggregation that is being used in volcanology (Aspinall 2010), by the Intergovernmental Panel for Climate Change (Kunreuther et al. 2014), and in several other applications of science. From an epistemological viewpoint, we should prefer judgment aggregation to singular thinking, even though aggregation has limitations and can be problematic (Martini and Sprenger 2017).

B) **Consider diversity in groups**. Diversity plays an important role in avoiding some of the biases that affect group decision-making and group-deliberation. Diversity ought to be limited when it affects the quality of the expertise base. To explain, there are two important elements that need to be considered when using expert judgment: the level of expertise and the diversity of the group. If the group lacks diversity, especially if the type of problems it deals with are complex and open-ended, then it is advisable to add diversity (Page 2008). But the diversity imperative cannot be absolute: If in order to add diversity we are adding group members that affect the level of expertise of the group, then we must be careful not to decrease the epistemic worth of the collective. In short, expertise and diversity must be balanced with one another.

6. Conclusion

In this article I have reviewed the role of expert judgment in epidemiological and pandemic modelling. I have highlighted how epidemiological and pandemic modelling are highly dependent on expert judgment, so much that getting the expertise right is as important as getting the model right. It is unlikely that there will be, in the short run, a highly accurate predictive model for pandemics; something like a descriptive model of celestial mechanics. If possible, that would be a welcome improvement, but for the time being we need to focus on expertise just as much as, if not more than, on technical issues about modelling.

Moreover, the starting and final points of modelling a pandemic are the same: expert judgment. That means that expert judgment is the first step in producing and parametrizing a model, and also that the product of a model of a pandemic is an input into the judgment of decision-making experts: "Above all, models should be viewed as aids to decision-making, rather than substitutes for decision-making" (Institute of Medicine 2006).

References

Achinstein, P. 1968, *Concepts of Science*, Baltimore: Johns Hopkins University Press.

Adam, D. 2020, "A Guide to R—The Pandemic's Misunderstood Metric: What the Reproduction Number Can and Can't Tell Us About Managing COVID-19", *Nature* 583, 346-48, doi: 10.1038/d41586-020-02009-w.

Aledort, J.E., Lurie, N., Wasserman, J., and Bozzette, S.A. 2007, "Non-Pharmaceutical Public Health Interventions for Pandemic Influenza: An Evaluation of the Evidence Base", *BMC Public Health*, 7, 208, doi: 10.1186/1471-2458-7-208.

Aspinall, W. 2010, "A Route to More Tractable Expert Advice", *Nature*, 463, 7279, 294-95.

Aspinall, W. and Cooke, R.M. 1998, "Expert Judgement and the Montserrat Volcano Eruption", in Mosleh, A. and Bari, R.A. (eds.) *Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management PSAM4*, September 13th -18th 1998, New York City, Vol. 3, London: Springer, 2113-18.

Benavides, A. 2018, "High-Altitude Forecasting: The Weather Wizards of the Greater Ranges", https://explorersweb.com/2018/11/20/high-altitude-forecasting-the-weather-wizards-of-the-greater-ranges/ (accessed: September 2021).

Bertozzi, A.L., Franco, E., Mohler, G., Short, M.B., and Sledge, D. 2020, "The Challenges of Modeling and Forecasting the Spread of COVID-19", *Proceedings of the National Academy of Sciences*, 117, 29, 16732-738.

Bishop, M.A. and Trout, J.D. 2005, *Epistemology and the Psychology of Human Judgment*, Oxford: Oxford University Press.

Black, M. 1962, *Models and Metaphors*, Ithaca: Cornell University Press.

Cleary, M., Lees, D., and Sayers, J. 2019, "Leadership, Thought Diversity, and the Influence of Groupthink", *Issues in Mental Health Nursing*, 40, 8, 731-33, doi: 10.1080/01612840.2019.1604050.

Cooke, R. 1991, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford: Oxford University Press.

Dreyfus, H.L. 1987, "From Socrates to Expert Systems: The Limits of Calculative Rationality", *Bulletin of the American Academy of Arts and Sciences*, 40, 4, 15-31.

Faust, D. 1984, *The Limits of Scientific Reasoning*, Minneapolis: University of Minnesota Press.

Goldberg, L.R. 1965, "Diagnosticians vs. Diagnostic Signs: The Diagnosis of Psychosis vs. Neurosis from the MMPI", *Psychological Monographs*, 79, 9, 1-28, doi: 10.1037/h0093885.

Häyry, M. 2021, "The COVID-19 Pandemic: A Month of Bioethics in Finland", *Cambridge Quarterly of Healthcare Ethics*, 30, 1, 114-22.

Kunreuther, H., Gupta, S., Bosetti, V., Cooke, R., Dutt, V., Ha-Duong, M., Held, H., Llanes-Regueiro, J., Patt, A., Shittu, E., and Weber, E. 2014, "Integrated Risk and Uncertainty Assessment of Climate Change Response Policies", in Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., and Minx, J.C. (eds.), *Climate Change 2014: Mitigation of Climate Change: Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge: Cambridge University Press.

Institute of Medicine 2006, "Modeling Community Containment for Pandemic Influenza: A Letter Report", Washington, DC: The National Academies Press, doi: 10.17226/11800.

Leli, D.A., and Filskov, S.B.F. 1984, "Clinical Detection of Intellectual Deterioration Associated with Brain Damage", *Journal of Clinical Psychology*, 40, 1435-41.

Mäki, U. 1992, "On the Method of Isolation in Economics", *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26, 4, 317-51.

Mäki, U. 2001, "Models, Metaphors, Narrative, and Rhetoric: Philosophical aspects", in Smelser, N.J. and Baltes, B. (eds.), *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam: Elsevier, 9931-37.

Martini, C. 2019, "The Epistemology of Expertise", in Fricker, M., Graham, P.J., Henderson, D., and Pedersen, N.J. (eds.), *The Routledge Handbook of Social Epistemology*, New York: Routledge, 115-22.

Martini, C. and Sprenger, J. 2017, "Opinion Aggregation and Individual Expertise", in Boyer-Kassem, T., Mayo-Wilson, C., and Weisberg, M. (eds.), *Scientific Collaboration and Collective Knowledge*, Oxford: Oxford University Press, 180-201.

Meehl, P.E. 1986, "Causes and Effects of my Disturbing Little Book", *Journal of Personality Assessment,* 50, 3, 370-75.

Mill, J.S. 1882 [1843], *A System of Logic, Ratiocinative and Inductive* (Eight Edition), New York: Harper & Brothers.

Molloy, J.C. 2011, "The Open Knowledge Foundation: Open Data Means Better Science", *PLoS Biol*, 9, 12, e1001195, doi: 10.1371/journal.pbio.1001195.

Morse, S.S., Garwin, R.L., and Olsiewski, P.J. 2006, "Next Flu Pandemic: What to Do Until the Vaccine Arrives?" *Science*, Nov. 10, 314, 5801, 929, doi: 10.1126/science.1135823.

Page, S. 2008, *The Difference*, Princeton: Princeton University Press.

Seixas, F.L., Zadrozny, B., Laks, J., Conci, A., and Saade, D.C.M. 2014, "A Bayesian Network Decision Model for Supporting the Diagnosis of Dementia, Alzheimer's Disease and Mild Cognitive Impairment", *Computers in Biology and Medicine*, 51, 140-58.

Samuel, O.W. and Omisore, M.O. 2013, "Hybrid Intelligent System for the Diagnosis of Typhoid Fever", *Journal of Computer Engineering and Information Technology*, 2, 2, doi: 10.4172/2324 9307.

Tversky, A. and Kahneman, D. 1974, "Judgment Under Uncertainty: Heuristics and Biases", *Science*, 185, 4157, 1124-31.

Walton, D., Reed, C., and Macagno, F. 2008, *Argumentation Schemes*, Cambridge: Cambridge University Press.

Grassroots Modeling during the Covid-19 Pandemic

Cecilia Nardini and Fridolin Gross***

** European School of Molecular Medicine (SEMM) and University of Milan*

*** ImmunoConcept, University of Bordeaux*

Abstract

One of the many peculiar phenomena that the Covid-19 pandemic has brought about is the engagement of non-scientists with specific questions surrounding the interpretation of epidemiological data and models. Many of them have even begun to get involved in the collection, analysis, and presentation of the data themselves. A reason for this might be that the insights that science can provide in a situation of crisis are often inconclusive or preliminary, motivating many people to look for the answers to pressing questions themselves. Moreover, public engagement is facilitated by the easy availability of up-to-date information, of the computational methods to process and analyze it, and of the infrastructure to share and communicate it with like-minded people. This raises epistemological questions about the status of such activities. Can they be considered scientific, and do they meet the standards of scientific inquiry? Or are they harmful because they add to the already loud chorus of voices spreading misinformation and increasing skepticism about the conventional scientific process? We propose to approach this question by looking at a concrete example: A community of active non-professionals has formed in Italy on the software development platform GitHub, where the Italian government's epidemiological data are made publicly available. This represents a well-defined and coherent case study on which detailed information is readily available.

Keywords: Citizen science, Data science, Covid-19, Pseudoscience.

1. Introduction

Nembro, in the province of Bergamo, is the municipality most affected by Covid-19 in relation to the population. We do not know exactly how many people have been infected, but we know that the number of deaths officially attributed to Covid-19 is 31. We are two physicists: one who became an entrepreneur in the health sector, the other a mayor, in close contact with a very cohesive territory, where we know each other very well. We noticed that something in these official numbers did not come back right, and we decided—together—to check.[1]

[1] Cancelli and Foresti 2020 (Claudio Cancelli, Mayor of Nembro, and Luca Foresti, founder of Centro Medico Santagostino).

Argumenta 7,1 (2021): 161-178
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.nar

The Covid-19 pandemic has changed the lives of people all over the world and has altered the way people work, meet, spend their free time, and get healed. And it has also, at least temporarily, changed scientific practice. It is undeniable that research has witnessed an enormous drive to produce results that could help understand the mechanism of transmission of the SARS-Cov2 virus, contain its spread and develop an effective vaccine, all much faster than would have been the case under normal circumstances. At the same time, the traditional scientific method has come under pressure: the conventional peer-review process has been struggling to keep up with the need for fast advancement; the rush to publish often leads to partial results or premature conclusions; and scientific claims are exploited in uncontrollable ways by politicians, the media or other individuals or institutions with an agenda.

In addition to these developments, there has been an unprecedented interest of the public in the details of scientific investigation. Given the direct relevance to their daily lives, people want to understand the numbers that are presented to them by the governments and the media, and to form an opinion on the way in which the pandemic is handled by the responsible institutions. But beyond the interest in existing information and analysis, we observe a significant interest from non-experts to participate in the process of data processing and analysis themselves. This kind of participation is facilitated by the fact that raw data is often publicly available and that it is now easy to obtain state-of-the-art computer tools for analysis and to share and discuss data and results online.

This raises the question whether, and to what extent, these kinds of activities can be considered 'scientific'. More specifically, one may ask to what extent these emerging structures resemble the organization and practices of professional science and whether they have the potential to lead to scientifically respectable outcomes. There is of course a risk that data analysis and modeling carried out outside the realm of conventional science may be used to spread misinformation or to contribute to the acceptance of harmful conspiracy theories. On the other hand, it seems that in the rapidly evolving situation of a global pandemic, conventional science cannot always provide interpretations and predictions quickly enough to meet the needs of the public. Instead of representing an alternative to conventional science, the efforts of non-professional modelers and data analysts may thus be understood as supporting and complementing science in relevant ways. It seems rather obvious to consider these activities as a form of 'citizen science', but at the same time there are clear differences to the paradigmatic examples of citizen science that have been discussed in the literature.

Instead of aiming at an exhaustive overview of the activities of non-professionals related to the pandemic, we decided to focus on a well-defined case study: the community of users of Covid-related data published by the Italian government on the software sharing platform GitHub. The structure of this platform has allowed us to easily follow discussions between members of the community, to track their modeling efforts and analyses, and to identify the outlets that they use to communicate their results to a wider audience. Moreover, the Italian context seems to be particularly interesting, as the situation there was very serious at the beginning of the pandemic, suggesting that the efforts of non-experts are not only driven by curiosity, but by a direct urge to contribute to and accelerate the management of the pandemic crisis.

The paper is structured as follows. In Section 2 we discuss the existing literature on citizen science and place it in the current context of the pandemic to provide the conceptual basis for framing our case study. We present this case study in detail in Section 3 and discuss it in Section 4. Section 5 offers concluding remarks.

2. Citizen Science

The phenomenon we wish to investigate consists in the increased participation of non-specialists in activities that bear similarities to scientific endeavors. Therefore, it is plausible to consider it within the framework of citizen science. In this section we discuss the way in which the concept of citizen science has been understood in the literature, and we motivate why the idea of citizen science gains particular relevance in the context of the current pandemic crisis.

The term 'citizen science' is relatively new, but it is often pointed out that before the professionalization of science in the 19th century, basically all science was citizen science (Cavalier and Kennedy 2016). The increased attention towards the end of the 20th century and the introduction of the label can thus partly be understood as a reaction to an increasing distance of science from the concerns of the public. The British sociologist Alan Irwin conceived of citizen science as a way of turning science into a more democratic endeavor (Irwin 1995). At roughly the same time, however, the term was also coined in a less politically charged way by Richard Bonney to describe the contribution of scientific data by nonscientists in the context of ornithology projects at Cornell University (Bonney 1996). In line with this, Cooper and Lewenstein (2016) distinguish between two meanings of citizen science: *democratized citizen science* and *contributory citizen science*. An example of democratized citizen science is the involvement of AIDS activists in scientific discussions in the mid-1980s to loosen restrictions on clinical trials and make newly developed treatments available to a wider audience. Prime examples of contributory citizen science are the activities of bird watchers or hobby astronomers who provide the results of their observations to scientific databases.

Contributory citizen science is typically more tightly integrated with science in the traditional sense: hobbyists and laypeople participate in data collection and other data intensive activities that are in turn built on by professional scientists to address relevant problems in their respective fields. However, the contribution of amateur scientists, valuable as it may be, is almost never original, creative, or critically aware. Democratized citizen science, by contrast, is situated at the interface between science and the public and may be seen as a form of interest group advocacy rather than as an epistemic endeavor, although there are cases where people substantiate their concerns by engaging in epistemic activities. An example is the Flint Water Study that involved citizens taking water samples to determine the lead concentration under the direction of professional scientists (Cooper and Lewenstein 2016). The contribution of this type of citizen scientist can be significant in advancing a particular line of research or raising real methodological questions. However, this also makes democratized citizen science more difficult to accept and evaluate this kind of citizen science as a genuine scientific activity.

The polysemous nature of the term makes it hard to find a unifying definition of 'citizen science'. A common theme, however, is that citizen science refers to activities that are carried out in direct interaction with professional scientists. Thus, democratized citizen science aims at convincing scientists of the importance of a particular cause and thereby to exert influence on the direction of research and on scientific policy making. Contributory citizen science, on the other hand, takes place in the context of projects that are created and supervised by professional scientists. In line with this, the Oxford English Dictionary defines 'citizen science' as "scientific work undertaken by members of the public, often in collaboration with or under the direction of professional scientists and scientific institutions".[2]

Citizen science has mostly been discussed from a sociological perspective, and it has not yet received much attention by philosophers. Existing philosophical discussions have mainly focused on the question whether the contributions of citizen science meet the standards of serious scientific inquiry. For example, Elliott and Rosenberg (2019) discuss three concerns about the quality of citizen science: that citizen science is not hypothesis-driven; that the collected data are of insufficient quality; and that citizen science is biased to the extent that it is politically motivated. They argue that none of these concerns threaten the potential value of citizen science. They point out, for example, that philosophers of science more generally have challenged the notion that all scientific activity must be guided by hypotheses. Thus, scientists themselves switch back and forth between different modes of research, some of which are purely exploratory or data driven. Overall, the activities of citizen scientists are presented as potentially valuable contributions to established science, either by directly adding to scientific projects, by directing scientists to issues of public concern, or by critiquing and modifying established scientific methods.

The Covid-19 pandemic raises pressing questions about the way in which science should be organized in a time of crisis. Answering such questions is not only important for dealing with the current situation, but may also be important in the longer term, as we can assume that similar global crises will occur even more frequently in the future. Philosophers have already given considerable thought to these issues,[3] and some of the most critical problems that have been raised can be understood as pointing to the importance and potential positive impact of citizen science, but also to the risks that such activities may entail in the current context.

First, there is the problem of urgency: science needs to react in a timely manner, and it needs to allocate its scarce resources in the best possible way to produce relevant and reliable results (Reydon 2020). Citizen science projects can help alleviate this problem by contributing to data collection or routine tasks that can be easily outsourced. There have been several examples illustrating the potential of such projects. The Eterna OpenVaccine project enables video game players to "design an mRNA encoding a potential vaccine against the novel coronavirus" (Do Soon and the Eterna Developer Team 2020). Another example is a project launched by UCSF via a smartphone app (Norris 2020), a remote public health study that collects data from participants on their habits and health status to gain insights into the spread of the virus. Lastly, the Rosetta@home (Peckham 2020) crowd-sourcing initiative harnesses the computational power of participants' home computers to find candidates for antiviral drugs. In this latter case the layperson will have the satisfaction of knowing they are contributing to scientific research, even without any original input on their part. Similar examples have

[2] https://www.oed.com/view/Entry/33513?rskey=skqsuT&result=1#eid316619123 (last accessed 07/11/2021).

[3] See the collection of short papers in a recent issue of HPLS, introduced by Boniolo and Onaga (2021).

been discussed in the literature, raising the question whether they constitute genuine cases of citizen science at all (Del Savio et al. 2016).

The second problem is the risk of science developing a "myopic, epidemiology-centric description of reality" (Lohse and Bschir 2020). In other words, there is concern that certain scientific disciplines, such as epidemiology or virology, are being given too much weight at the expense of other relevant fields and perspectives. In response to this, many philosophers have emphasized the need for a more pluralistic approach to the processes of knowledge generation and policy-making that should involve as many stakeholders as possible (Mazzocchi 2021; Ongaro 2021; Leonelli 2021). Clearly, forms of democratized citizen science are one way to address this problem of lack of pluralism, for example by focusing scientists' attention on relevant local contexts or particularly affected population groups. For example, one activist group has written an open letter urging the NIH to include patients with HIV/AIDS in trials of the new SARS/Cov2 vaccines.[4]

Finally, there is the problem of uncertainty and misinformation. Faced with incomplete knowledge and uncertain evidence, scientists have openly disagreed about the best ways to deal with the pandemic,[5] and the accelerated scientific process has led to misuses of results and to the retraction of findings. As a result, large parts of the public have lost trust in the scientific process, which in turn plays in the hands of denialists who question the seriousness of the problem and the need for action to combat the virus (Antiochou 2021; Monasterio Astobiza 2021). Differently from the others, this issue makes the potential role and value of citizen science seem rather ambivalent. On the one hand, citizen science initiatives might have a beneficial impact by critically assessing the way in which science is done, thereby achieving increased transparency and public understanding. On the other hand, there are obvious risks that these activities may lead to the spread of misinformation, adding yet another voice to the already loud chorus that undermines the credibility of conventional science.

In what follows we would like to illuminate these problems and the potential role of citizen science using a concrete case study. This case study shares important similarities with the two types of citizen science identified at the beginning, as it clearly involves a topic of public interest while also mobilizing the skills of many amateur data scientists. At the same time, it seems very different because it looks like a largely self-organized 'grassroots' effort by non-scientists that is not directly linked to the Covid-related projects of the scientific community. Given these features, we think that our case study can contribute to a better understanding of both the value and the potential risks of public participation in the process of knowledge generation and interpretation of scientific evidence. More specifically, we would like to understand whether such activities, when conducted largely independently from established science, necessarily fall into the camp of 'pseudoscience,' whether they lack the quality that other citizen science projects have because of support from professional scientists, or whether they can be understood more positively as indicative of an alternative, more open and inclusive model of scientific research.

[4] https://www.treatmentactiongroup.org/wp-content/uploads/2020/08/covid_19_1273_collins_nih_7_27_20.pdf (last accessed 07/11/2021).

[5] For an example consider the exchange between Ioannidis and Lipsitch (Ioannidis 2020; Lipsitch 2020).

3. The Case Study

On 21 February 2020, ten small municipalities in Lombardy were quarantined after the discovery of a local hotspot of the new Coronavirus disease. In the following weeks, the pandemic took the country by storm. Italy was the first western country to feel the effects of the new virus and the first to enter a nationwide lockdown on 9 March. The Italian Civil Protection Department (DPC) was engaged since the beginning of the emergency and started releasing daily communications of the epidemiological situation by assembling data from all the different regional health systems. These reports were made public every day at 6PM in a press conference and subsequently published in a .pdf document on the official website of the DPC. On 4 March 2020, the open data association onData[6] began automatically scraping these files to publish the same data on the software development platform GitHub in a format more suitable for further computational analysis[7], while petitioning the Italian government to publish the data in an open, machine-readable format directly from the DPC. On 7 March 2020, the DPC opened its own GitHub repository, where the data contained in the daily report were published daily after the press conference in machine-readable format and under a Creative Commons license. Since 25 June 2020, the data has been released directly by the Ministry of Health, but the open data repository continues to be curated by the DPC.

A community of users quickly formed around both the first unofficial repository and the official DPC repository, using the data for a variety of purposes, such as personal or publicly available spreadsheets or dashboards to monitor the progression of the pandemic. The GitHub 'Issues' system provides a convenient way to map and analyze this community. 'Issues' are online discussions usually related to queries, clarification requests or bug reports that can be opened on a repository by registered users. Although they are mainly intended as a means of communication between the maintainers and the users of the repository, issues can—and often do—become a place of communication and exchange among a broader community of users, especially when the repository is public. This has happened to a great extent with the 'Issues' section of the DPC repository: with several issues opened every week and only two official maintainers actively taking care of the DPC repository—users @umbros (Umberto Rosini) and @pierluigicara (Pierluigi Cara)—queries and clarifications from external users are often answered by other external users, leading in some cases to long and intense discussions.

3.1 Methods

To analyze the community of users of the GitHub repository of the Italian Government, we surveyed the approximately one thousand issues published there from the date of its creation (7 March 2020) to April 2021 and selected those that met our criteria for interest/significance.

Issues generally fall into two broad categories. A subset of issues are opened to signal minor inconsistencies or errors in the data that the users can quickly spot because of the type of automated analysis they perform on the data. Issues of this

[6] https://ondata.it/ (last accessed 06/11/2021).

[7] https://github.com/ondata/covid19italia (last accessed 06/11/2021).

type usually receive a reply from the maintainers of the repository and are sometimes passed on to the DPC/Health Ministry to prompt a correction.

We chose to focus on a second, more philosophically interesting kind of issues that contain open-ended, methodological discussions. Here, maintainers intervene sparingly, if at all, while the participation of other members of the community is often lively. Furthermore, the questions posed in these issues often remain unresolved, which provides an interesting parallel to open research questions in traditional science.

After identifying the most interesting issues in this way, we tracked the users participating in the discussions to identify possible modeling and analysis efforts beyond their contribution to the repository. This was easy when the users published their work in their own GitHub repository or on a personal website linked to their GitHub profile. In some cases, however, it was impossible to track down this additional work from a user's GitHub profile, even if they had mentioned it in the discussion.

In the following, we will first present the individual issues that we believe are relevant for the context of our paper. Afterwards, we will present four examples of larger projects and several examples of individual users who engaged in modeling based on the data provided by the repository.

3.2 Findings

3.2.1 Individual Issues

Our first step in analyzing the community of GitHub users consisted in a survey of individual issues in the repository.

Of all the topics we captured for follow-up and used to identify the users' modeling efforts, we describe below a selection of issues that we believe can provide a representative overview of the type of discussions taking place among community members.

- Issue #577 concerns the data collected in the field 'tamponi' (swab tests) of the published data. The discussion clarifies that in different Italian regions data are generated differently. For example, in some regions they include all swabs, while in others only swabs that have already been analyzed are counted.
- Issue #587 concerns the estimation of R_0 and R_t, the initial and the current reproduction number of the virus. This is a very interesting issue because several of the participants propose their own analysis of these metrics. For example, users @alessandroNa, @Riccardocominotti, @LucaZeta, @brunocaniglia and @mpreitano present their methods for a simplified estimation of R_0, and they suggest possible improvements to each other. Additionally, user @Pivone presents his detailed analysis based on several indicators constructed from the DPC data (this user will be treated more in depth in Section 3.2.3).
- Issue #821 concerns two new fields that were added to the dataset (and later removed): 'Casi da sospetto diagnostico' and 'Casi da screening' (cases found due to diagnostic suspicion or via screening, respectively). Participants in the discussion debate the correct interpretation of the two new fields and provide evidence that the definitions of the two metrics are interpreted differently by different regions, leading to inconsistencies in the data. The distinction is considered relevant because of the different probabilities of finding an active (contagious) case by the two different methods. In the

discussion, two different interpretations are proposed: according to users @Paulsword and @Rabelaiss the first category (diagnostic suspicion) includes also cases found via contact tracing, so that cases in this category are more likely to be active spreaders of the infection. According to user @vienne, by contrast, contact tracing falls in the second category (screening), which is therefore the category that has the greater likelihood of including infection hotspots. Both sides of the discussion support their point of view by citing the national or regional health authorities and data. Ultimately, however, the issue remains unresolved, as further clarification from the Ministry of Health is still pending at the time of writing, and the new fields are removed from the dataset anyway in a later revision.

- Issue #864 features a very interesting and long discussion on the definition of some of the main quantities provided as open data on the repository, the fields 'Casi testati' (people tested) and 'Tamponi' (samples tested). These definitions are crucial because the two measures are used by health authorities and the media as a basis to calculate the daily incidence figures. In the discussion it is noted how ambiguity or inconsistency in definition may lead to systematic overestimation or underestimation of the daily incidence. The arguments put forward by the participants are valid, but they leave open the question of whether the competent health authorities have made the same considerations.
- Issue#892 highlights an apparent inconsistency in the trend of the number of recovered patients vs new cases. The claim is backed by a graphical analysis of the data in question. However, there is no official acknowledgment of the anomaly.
- Issue #977 is a very debated one in which at least two interesting questions are analyzed. The starting point is the introduction of a new field in the dataset ('Ingressi del giorno in terapia intensiva', daily new entries in ICU), and participants discuss the relation between this new field and other quantities in the dataset. A second question that appears in the same issue is the usefulness of an index, introduced by user @CT-igiul, based on the relative variation of current positive cases. User @Rabelaiss argues that this index does not give a useful picture of the evolution of the pandemic, while user @Doc73 points out that it bears similarities with the technique of derivative control used in industrial control systems. This issue is interesting in terms of its content, but also because it represents an example of methodological discussion in which the community productively engages with the work of one of its members.
- Issue #1005 concerns the observation of suspicious simultaneous spikes in weekly averages of deaths, cases, and tests. Some explanatory hypotheses are proposed, but again there is no official acknowledgment of the problem.
- Issue #1136 is opened by user @AntonioB1976 as a fact-checking request into a Covid-19 denialist's claims on Facebook that the number of new positive cases reported daily by authorities includes repeated tests of already know positive cases. None of the maintainers intervene to make an official statement, but some of the most active users provide data and observations to refute the denialist's claims.

The issues we have singled out constitute a representative sample of the kind of interaction and dialogue that takes place in this community. As can be seen, the general tone is altogether different from other social media forums. Issues are usually opened with a precise methodological or data-related question in mind, and the answers are not purely opinion-based, but are usually supported with references to scientific literature, to health authorities or directly with data and analysis results. The debates are rational, and the common goal appears to be that of gaining a better understanding of the underlying issue or of the data, rather than to convince others of one's own opinion. On the other hand, the discussions resemble those on other social forums in that their impact is limited to the participants or, at best, to other interested members of the community. In some instances (e.g., in issue #821 mentioned above) official maintainer @umbros intervened to say that he would submit a query to the Health Ministry for clarification, but in all cases were this was done, the official response, if ever given, was not reported back on the repository.

3.2.2 Larger Projects

After looking at individual issues, we proceeded to track participants through their GitHub profiles and assessed whether there was any research projects available on their GitHub profile or otherwise reachable via links from there. We were able to identify several web dashboards fed with the DPC data from the repository, and we included them in our analysis if there was evidence of original research content beyond mere reporting or visualization of the data. In the following we will detail the main modeling efforts that we identified in this way.

EpiDataItalia

According to their website,[8] EpiDataItalia is a study and research group formed by three self-funded volunteers (a data scientist, a biologist by training and a musician who is an amateur mathematician/statistician). It seeks to offer analysis and forecast on the COVID-19 pandemic with particular attention to the situation in Italy. Apart from directly posting on their website, they have published their results and analyses in the news magazine *L'Espresso* and as preprints on open access repositories, such as *Zenodo*.

One part of their project consists in processing and visualizing the data provided at the national level by the Italian government and at the international level by Johns Hopkins University. However, they go beyond mere reporting of data by pursuing their own scientific questions, for example the correlation between air pollutants and COVID-19 cases in the Lombardy region. They also compared in detail different methods for calculating the effective reproductive number R_t, proposed a new way of estimating the case fatality rate, and investigated the consequences of different vaccination strategies using mathematical models.

ilsegnalatore.info

This is mostly a scientific news website, originally created to provide controlled, verified pieces of news and information on the pandemic. The main author of the site is a physician, Paolo Spada (user @paulsword on GitHub), who is an active participant in many of the GitHub issues that we analyzed. Since March 2020 he has published a detailed daily report with infographics on the DPC data. The report, published both on the website and on his own Facebook page, is followed by thousands of citizens and has attracted national-level attention with an

[8] https://www.epidata.it last accessed 06/11/2021.

interview in the magazine *Panorama* (Bonaccorso 2020) and several interviews on national television.[9]

The detailed daily report is interesting because it contains elaborations that go beyond the mere visualization of the time series of data. For instance, in a graph recently added to the report, trends in incidence rates, rates of ICU admission and fatality are plotted against vaccination coverage in the 60+ age group, with the assumption that the latter two values should be lower than in the previous waves because of the protection offered by vaccination to the most vulnerable age group.

In addition to the daily reports, Spada has published around 20 articles on the website. Some are mainly scientific communication articles: for example, there is an explanation and commentary on a graph published in *JAMA* depicting how the probability of detecting an infection with different tests varies over time,[10] and an explanation of the meaning of the various indicators that are communicated daily by newspapers. However, some go beyond scientific communication by including a critical review of the available data, often in the light of current scientific literature. For instance, in an early article[11] he compared the predictions of the *SIR* model with the actual data to show an apparent overestimation of recovered patients in the data reported by the Lombardy region. Another one, "Oltre l'R_t",[12] ("Beyond R_t"), proposes and evaluates the use of the weekly average of the percentage variation of new cases as a proxy measure for R_t. The interest of this proxy measure is that it is a value that is readily available from the data up to the current day, unlike R_t itself which has a considerable lag because it needs to consider the time interval between the infection and the onset of symptoms and between the onset of symptoms and the diagnosis. The comparison between the two values (weekly average of the percent variation vs. R_t shifted back in time) is shown daily in the reports on the website and there is indeed a strong correspondence between the two curves.

OpenCovidItaly initiative

OpenCovidItaly is one of the data users that we identified starting from the first unofficial OnData repository. It is a blog/study group that published several articles and analyses between May and August 2020.

There is neither a detailed description of the group's structure nor a listing of the individual participants, but the "Perché" (Why) section of the blog[13] explains that the main motivation of the initiative is to provide open data on the pandemic through scraping and collecting data from various sources.

The first posts are indeed just data presentation, providing a breakdown of the data about deaths in some Italian regions. However, subsequent posts include some elaborations on the presented data. In particular, there is a methodological

[9] For instance, https://ilsegnalatore.info/frontiere-raitre/ or https://ilsegnalatore.info/tg5-ore-20-mediaset-4/ (last accessed 06/11/2021).

[10] https://ilsegnalatore.info/una-figura-e-meglio-di-tante-parole/ (last accessed 06/11/2021).

[11] https://ilsegnalatore.info/i-pazienti-dimenticati-nei-conti-della-lombardia/ (last accessed 06/11/2021).

[12] https://ilsegnalatore.info/oltre-allrt/ last accessed (06/11/2021).

[13] https://opencoviditaly.netsons.org/why/ (last accessed 06/11/2021).

article[14] explaining the risk of using data that are not yet consolidated, and an explanation[15] of the epidemiological indicator R_t with an in-depth analysis of how it can be estimated from data that are constantly under accrual.

Currently, they use their twitter profile[16] to publish a weekly forecast of the value of R_t. They then proceed to confront this forecast with the official figure released by the National Health Institute the following day. This is a particularly interesting example of the interaction of grassroots modelers because this forecast for R_t is obtained using a web application published by another participant in the community (user @vi-enne mentioned below), fed with data provided by the OnData collective, the original unofficial source of data which currently is still providing some finer-granularity data that would otherwise be unavailable in machine-readable format.

Vittorio Nicoletta

User @vi-enne (Vittorio Nicoletta) is a computer scientist active on twitter with the handle @vi__enne.[17] He has a public repository[18] in which he answers some frequently asked questions on Covid with explanations, data, and literature references. He also published an analysis dashboard (a public Google Drive worksheet) for forecasting the level of risk and transmission in the different Italian regions. Finally, he created a web application[19] that allows any web user to estimate the value of R_t from the official data or from any dataset they provide in .csv format. The app uses the models available in the EpiEstim R software package and allows the user to set some analysis parameters and choose between the four available models for estimation. This is the application mentioned above that is used by the OpenCovidItaly group to estimate their weekly R_t forecast.

3.2.3 Others

Besides the more prominent examples that we have mentioned so far, we have identified several less systematic but still noteworthy modeling efforts.

- GitHub user @alexamici (Alessandro Amici) has a repository[20] of Python notebooks that are updated daily with data and short-term forecasts at national and regional levels. He also has a blog[21] that he updated between March and October 2020 with statistical and data analytic considerations.
- Users @littleark (Carlo Zapponi) and @leeppolis (Simone Lippoli) are the founders of *Visualize News*, a group of computational designers with an interest in data visualization. They curate an infographics dashboard on Covid[22] which also includes some elements of original analysis, e.g., the section “How is today’s situation compared with the first wave?”

[14] https://opencoviditaly.netsons.org/cosa-succede-quando-si-utilizzano-dati-non-consolidati/ (last accessed 06/11/2021).

[15] https://opencoviditaly.netsons.org/erreti-leggermente-maggiore-di-uno/ (last accessed 06/11/2021).

[16] https://twitter.com/OpencovidM (last accessed 06/11/2021).

[17] https://twitter.com/vi__enne (last accessed 06/11/2021).

[18] https://github.com/vi-enne/FAQ_covid19_ITA (last accessed 06/11/2021).

[19] https://vienne.shinyapps.io/rt_estimation/ (last accessed 06/11/2021).

[20] https://github.com/alexamici/covid-19-notebooks (last accessed 06/11/2021).

[21] https://naturalstupidity.ghost.io (last accessed 06/11/2021).

[22] https://coronavirus.visualize.news/ (last accessed 31/05/2021).

- User @fotografAle (who is, according to his profile, a professional photographer) was active in some of the issues analyzed above. In one of them he attached a plot and referenced an analysis that resulted from a deep learning forecast model he created. Unfortunately, the model (which he says is based on a convolutional neural network) is not published in his GitHub profile and does not appear to be publicly accessible, so this mention in one of the issues is the only reference to its existence.
- User @heyteacher has a GitHub repository[23] with a machine learning project for forecasting the evolution of the pandemic. Unfortunately, it appears to be abandoned (last updated in June 2020), and there is no way to assess it now. The dashboard[24] by the same author is still updated but contains only data visualization.
- User @LucaZeta is another very active participants in many of the issues. He has created a dashboard[25] that looks quite confusing. There are lines in the graph labeled as 'analysis' but no indication at how they are derived.
- User @vitop72 has a public repository named Covid19 Italy Report[26] updated between April and June 2020 with weekly reports (in .pdf format) that contain a graphical elaboration of the various pandemic-related indicators and a forecast for the coming week.
- User @CT-igiul (Luigi Tomaselli) was very active in some of the issues. He publishes a blog[27], still updated as of May 2021, with some analyses and elaboration; in particular, he has developed an indicator based on the daily relative variation of current positive cases.
- User @Pivone participated in one of the issues posting details and some results of his analysis.[28] He developed some indicators for the development of the pandemic, such as a simple estimate of R_0 and a linear regression. He also analyzed the ratio between home quarantined patients and patients in hospitals for various regions. This allowed him to conclude that in the first months of the pandemic in Lombardy mostly only people with severe symptoms were tested, a fact that has since been officially recognized.
- User @SilvioCaggia (Silvio Caggia) also shared his analysis in the context of an issue.[29] His model is a Google Drive spreadsheet document[30] with graphs and visualizations of the DPC data, but there are hints of original analysis, for instance the sheet 'Qcomparativo' which, as he writes in the issue, is an attempt to analyze fatality rates in different regions.

All the models that we examined can be placed on an axis where, on one end, there are personal research efforts that users pursue in isolation and are reluctant to share (e.g. the projects of users @fotografAle, @Pivone and

[23] https://github.com/heyteacher/sam-forecast-automation-covid-19-ita (last accessed 06/11/2021).

[24] https://heyteacher.github.io/COVID-19/#/ (last accessed 06/11/2021).

[25] https://covid19.zappi.me/ (last accessed 06/11/2021).

[26] https://github.com/vitop72/Covid19-Italy-Report (last accessed 06/11/2021).

[27] https://www.luigitomaselli.com/ (last accessed 06/11/2021).

[28] https://github.com/pcm-dpc/COVID-19/issues/587#issuecomment-637168807 (last accessed 06/11/2021).

[29] https://github.com/pcm-dpc/COVID-19/issues/759 (last accessed 06/11/2021).

[30]lhttps://docs.google.com/spreadsheets/d/11S6KS8lpYq_rNYdf4uqZhKgmPSmHvG9s7S0O-dD4PH4/edit#gid=1092157180 (last accessed 31/05/2021).

@SilvioCaggia described in Section 3.2.3), while, on the other end, there are public dashboards or blogs whose main motivation is scientifically supported public communication (the cases of Visualize News and ilsegnalatore.info are the most obvious examples).

Between these two extremes, some projects are shared with a less wide audience in mind, that is, with a community of experts and insiders. This is the case, for instance, of the application developed by Vittorio Nicoletta for the estimation of R_t, of many of the elaborations by the OpenCovidItaly initiative, or of the blog curated by user @alexamici. For these kinds of projects, the main distribution channels outside of GitHub are traditional social media, such as Twitter. Indeed, from a brief analysis of this informal 'citation network' we found a certain level of interplay between these projects.

In the next section we will consider what our findings entail for the original questions we set out to examine.

4. Discussion

Our case study provides detailed insight into a community of non-professionals who engage in the presentation and analysis of data related to the current pandemic. How seriously should one take this kind of activity? Can it be called 'scientific', or is it more the hobbyhorse of a group of 'data nerds' who play scientists to entertain themselves while they are stuck at home? While it is difficult for us to directly assess the scientific merit of the analyses and models proposed by the members of the community, we can at least look at their practices and interactions as revealed in the GitHub discussions and compare them to genuine scientific activities.

There are several ways in which what we observe resembles the practices of professional scientists (or at least the normative ideal of science). First, the discussions are constructive and rational and very different in style from the kinds of discussions one can witness on other social media platforms. Participants usually share a common goal of better understanding a particular phenomenon, concept, or methodological issue, e.g., how well quantitative measurements provided by public institutions reflect the true dynamics of the pandemic. And they do not appear to be pursuing their activities for financial or personal gain. Second, the members of the community engage in open sharing of their results, resources, methods, and concepts that they use. This kind of collaboration is facilitated by the fact that data and software code are typically made openly available on GitHub, and we observe that in some cases participants use other participants' results for their own projects and build on them. Taken together, this suggests that members of this community adhere to the norms typically associated with the ideal of scientific conduct (Merton 1942; Anderson et al. 2010). Furthermore, we can find a certain degree of continuity between the activities of the community and established scientific research, in that the members of the community explicitly refer to scientific resources and make use of concepts and statistical tools from epidemiological research (e.g., by using the same software packages that are also used by professional scientists). And it does not appear that their work advocates doctrines that are in tension with established beliefs of the relevant scientific fields. All this suggests that the activities of the community cannot simply be dismissed as pseudoscientific in the same way as, for example, the alternative theories and models of climate change deniers (Hansson 2017).

On the other hand, there are clear differences between what we observe in the GitHub community and established science, for better or for worse. Overall, the activities of the community are less coherent, as everyone works mainly on their own problems and analytical tools, despite intense discussions and occasional sharing of resources. Moreover, there are no agreed standards or measures for peer control. Instead, users present their work to others and allow it to be critiqued on a purely voluntary basis. Finally, there are no restrictions on entry into the discussion. Participation in conventional science is typically restricted to individuals who have an accepted degree of qualification (e.g., a PhD) and are affiliated with an official institution (e.g., a university). The GitHub community, by contrast, is in principle open to anyone who has an account. Such openness carries an obvious risk of lowering the quality of the output of the community. However, this feature can also be seen positively as it increases the diversity of participants and can make research more productive and balanced, which, as we have discussed in Section 2, is especially relevant in the context of a social crisis such as a pandemic.

We think that the two standard modes of citizen science, democratized citizen science and contributory citizen science, do not really capture what we observe in our case study. First, the activities of the GitHub community are completely bottom-up and self-contained, i.e., carried out without any direct involvement of professional scientists. Secondly, the citizen scientists in our case study do not want to influence the scientists, but to take matters into their own hands: the community members would not be content to work within the framework of established methods, as part of their activities is precisely to question and criticize these methods.

Of the citizen scientists we encountered in our study, many are motivated by a desire to improve their personal understanding of the situation by analyzing the data on their own. Some users are skeptical of the way valid scientific results are interpreted by the media and disseminated to the public, and therefore develop their own measures and analyses to understand and critically evaluate the news. The words of user @Pivone, mentioned in Section 3.2.3, exemplify this attitude: "credo che questo set di dati aggregati [...] permetta di fare uno screening ragionato e serio delle notizie e di rigettare le molte cose fuorvianti che sono state dette e scritte in proposito"[31] (I believe that this dataset allows for a reasonable and serious screening of the news and to disprove the many misleading things that have been said on the matter).

In other cases, however, there is real dissatisfaction among modelers because they feel that conventional science and government agencies are overwhelmed and cannot respond with the necessary care or transparency in their communications. An example is the recurring issue of the right way to determine R_t, the reproduction number of the virus. There is no consensus on how best to estimate this number, but it appears to be crucial because it expresses in a simple way where the pandemic is headed. Thus, the members of the community respond to the need to increase transparency around this issue, feeling that the scientific community and government institutions at various levels often send conflicting messages. For example, user @PaulSword developed an alternative measure of the progress of the pandemic in his articles on ilsegnalatore.info, mentioned in

[31] https://github.com/pcm-dpc/COVID-19/issues/587#issuecomment-642287928 (last accessed 06/11/2021).

Section 3.2.2, based on weekly average variation of new cases. This metric, while being easy to compute and based on current data, provides a very good approximation of the official estimate of R_t. The good fit between the two measures is shown on the ilsegnaltore.info blog with weekly updated charts. His question as to why the authorities do not take this simplified model into account seems legitimate, especially since the measure he developed has the advantage of being almost real-time, unlike the official estimate, which can only be calculated with a delay of two weeks.

More generally, the focus on R_t, and on similar metrics that capture the progression of the pandemic offers important insight into the ultimately social motivations behind the research efforts of the community. Fostering collaborative and safe behavior in citizens through clear and accurate scientific communication is a strong motivation behind the effort of some of the modelers we studied. As pointed out earlier, this *bona fide* sentiment seems far removed from denialist positions or attempts to propagate conspiracy theories. Most members of the community appear to have professional experience in dealing with complex data and are therefore aware of the methodological pitfalls that can affect data-based decision-making in situations such as this one, where a large amount of heterogeneous data must be collected quickly, and the data collection pipeline had to be set up hastily and with little oversight.

An example that supports this idea comes from an altogether different case: the COVID Tracking Project in the U.S.,[32] which is a data collection initiative that was launched by the news magazine *The Atlantic* in March 2020 out of dissatisfaction with the data the U.S. Centers for Disease Control and Prevention (CDC) were making publicly available (Meyer and Madrigal 2021). Over a few months the project, which was based on the data collection effort by hundreds of citizen volunteers, became the most complete data source about COVID-19 in the U.S., being used by *The New York Times*, Johns Hopkins University, and two presidential administrations, as well as being cited in more than 1,000 academic papers, including major medical journals like *The New England Journal of Medicine*, *Nature*, and *JAMA*. The project's founders, Meyer and Madrigal (two journalists), emphasize the importance of understanding the way data is collected and organized in order to be interpreted correctly:

> The scientists at the CDC clearly have far more expertise in infectious-disease containment than almost anyone at the COVID Tracking Project or *The Atlantic*. But we did spend a year grappling with the limitations of the system (Meyer and Madrigal 2021).

The example of the COVID Tracking Project is, we believe, a success story that provides a glimpse of what the GitHub community could have looked like if somebody with the necessary power had managed to coordinate the efforts, the data skills, and methodological expertise of the participants.

For it must be acknowledged that while the members of the GitHub community have clearly achieved useful results, their activities remain somewhat fragmented and do not seem to be having the kind of impact that a better organized and streamlined project could have achieved.

[32] https://covidtracking.com/ (last accessed 06/11/2021).

5. Conclusion

In this article we examined the activity around the GitHub repository where the Italian government publishes data related to the Covid19 pandemic. What we have discovered is a peculiar kind of citizen science in which lay people try to improve the kind of information they get from official sources. We have argued that these activities can at least partly be considered scientific, but they are different from other forms of citizen science because they do not rely on the direct involvement of professional scientists. Rather, we observe that the citizen scientists in our case study attempt to circumvent or 'short-circuit' the usual flow of information that reaches the public via science or the media.

Obviously, our analysis provides only a short glimpse into a phenomenon that might itself be transient and dependent on the dynamics of the pandemic. Furthermore, we were not able to systematically track many of the activities of community members that took place outside the GitHub platform. Thus, it may be promising to map the informal 'citation networks' of citizen scientists across social media platforms such as Facebook and Twitter and compare them to the organization of established science.

Despite its epistemic shortcomings, we see the community described in our case study as a positive example that avoids some of the risks typically associated with public participation in controversial scientific topics, while at the same time exhibiting greater openness and diversity, a feature that seems particularly relevant in the current crisis.

We started our article with a quote from Claudio Cancelli, the Mayor of Nembro, and his fellow citizen scientist Luca Foresti. In the same article they sum up the particular requirements of the current situation:

> We are in the midst of an epoch-making event and to fight it we need credible data on the reality of the situation, disclosed transparently among all the experts and people who have to manage the crisis responsibly. Based on these data we can understand and decide what is right to do when it is required (Cancelli and Foresti 2020).

This call should be understood in its broadest sense, as we are all citizens involved in the responsible management of this crisis. Therefore, we believe it reflects the unprecedented momentum that has led many citizen scientists to commit their time and efforts to contribute to a better understanding of the Covid-19 pandemic.

References

Anderson, M.S., Ronning, E.A., Vries, R.D., and Martinson, B.C. 2010, "Extending the Mertonian Norms: Scientists' Subscription to Norms of Research", *The Journal of Higher Education*, 81, 3, 366-93.

Antiochou, K. 2021, "Science Communication: Challenges and Dilemmas in the Age of COVID-19", *History and Philosophy of the Life Sciences*, 43, 3, 87.

Bonaccorso, M. 2020, "Il Medico dei Numeri del Covid-19", *Panorama*, Mar 31, https://www.panorama.it/news/salute/il-medico-dei-numeri-del-covid-19 (last accessed 07/11/21).

Boniolo, G. and L. Onaga 2021, "Seeing Clearly through COVID-19: Current and Future Questions for the History and Philosophy of the Life Sciences", *History and Philosophy of the Life Sciences*, 43, 2, 83.

Bonney, R. 1996, "Citizen Science: A Lab Tradition", *Living Bird*, 15, 4, 7-15.

Cancelli, C. and Foresti, L. 2020, "The Real Death Toll for Covid-19 is at least 4 Times the Official Numbers", *Il Corriere della Sera*, English version, March 26, https://www.corriere.it/politica/20_marzo_26/the-real-death-toll-for-covid-19-is-at-least-4-times-the-official-numbers-b5af0edc-6eeb-11ea-925b-a0c3cdbe1130.shtml (last accessed 30/11/2021).

Cavalier, D. and Kennedy, E.B. (eds.) 2016, *The Rightful Place of Science: Citizen Science*, Tempe: Consortium for Science, Policy & Outcomes.

Cooper, C.B. and Lewenstein, B.W. 2016, "Two Meanings of Citizen Science", in Cavalier and Kennedy, 51-61.

Del Savio, L., Prainsack, B., and Buyx, A. 2016, "Crowdsourcing the Human Gut: Is Crowdsourcing Also 'Citizen Science'?", *Journal of Science Communication*, 15, 3, A03.

Do Soon and the Eterna Developer Team 2020, Eterna OpenVaccine, https://eternagame.org/ (last accessed 06/11/2021).

Elliott, K.C. and Rosenberg, J. 2019, "Philosophical Foundations for Citizen Science", *Citizen Science: Theory and Practice*, 4, 1, 9.

Hansson, S.O. 2017, "Science Denial as a Form of Pseudoscience", *Studies in History and Philosophy of Science*, Part A 63, 39-47.

Ioannidis, J.P.A. 2020, "The Totality of the Evidence", *Boston Review*, 26, 22-30.

Irwin, A. 1995, *Citizen Science: A Study of People, Expertise and Sustainable Development*, London: Routledge.

Leonelli, S. 2021, "Data Science in Times of Pan(dem)ic", *Harvard Data Science Review*, 3, 1, doi: 10.1162/99608f92.fbb1bdd6 (last accessed 28/11/21).

Lipsitch, M. 2020, "Good Science Is Good Science", *European Journal of Epidemiology*, 35, 519-22.

Lohse, S. and Bschir, K. 2020, "The COVID-19 Pandemic: a Case for Epistemic Pluralism in Public Health Policy", *History and Philosophy of the Life Sciences*, 42, 4, 58.

Mazzocchi, F. 2021, "Drawing Lessons from the COVID-19 Pandemic: Science and Epistemic Humility Should go Together", *History and Philosophy of the Life Sciences*, 43, 3, 92.

Merton, R.K. 1942, "A Note on Science and Democracy", *Journal of Legal and Political Sociology*, 1, 115-26.

Meyer, R. and Madrigal, A.C. 2021, "Why the Pandemic Experts Failed: We're still Thinking about Pandemic Data in the Wrong Ways", *The Atlantic*, March 15, https://www.theatlantic.com/science/archive/2021/03/americas-coronavirus-catastrophe-began-with-data/618287/ (last accessed 07/11/21).

Monasterio Astobiza, A. 2021, "Science, Misinformation and Digital Technology during the Covid-19 Pandemic", *History and Philosophy of the Life Sciences*, 43, 2, 68.

Norris, J. 2020, "New COVID-19 'Citizen Science' Initiative Lets any Adult with a Smartphone Help to Fight Coronavirus", *UCFS news*, March 30, https://www.ucsf.edu/news/2020/03/417026/new-covid-19-citizen-science-initiative-lets-any-adult-smartphone-help-fight (last accessed 07/11/21).

Ongaro, M. 2021, “Making Policy Decisions under Plural Uncertainty: Responding to the COVID-19 Pandemic”, *History and Philosophy of the Life Sciences*, 43, 2, 56.

Peckham, O. 2020, “Rosetta@home Rallies a Legion of Computers Against the Coronavirus”, HPCwire, March 24, https://www.hpcwire.com/2020/03/24/rosetta-home-rallies-a-legion-of-computers-against-the-coronavirus/ (last accessed 07/11/21).

Reydon, T.A.C. 2020, “How Can Science Be Well-Ordered in Times of Crisis? Learning from the SARS-CoV-2 Pandemic”, *History and Philosophy of the Life Sciences*, 42, 4, 53.

Science, Scientism, and the Disunity of Science: Popular Science during the COVID-19 Pandemic

Nicolò Gaj and Giuseppe Lo Dico

Catholic University of the Sacred Heart, Milan

Abstract

Unsurprisingly, science has been conferred growing expectations in the context of the COVID-19 pandemic. Accordingly, the issue of dissemination and popularization of scientific outcomes has come to the fore. The article describes the main features of the so-called dominant view in popular science, which is claimed to be implicitly connected to scientism, a stance identifying science as the most (if not the only) reliable source of legitimate knowledge. Scientism's implicit philosophical roots are argued to lie in naturalism and a trivialized neopositivist concept of science, which underscores the supposed unity of the scientific enterprise. However, in the context of the pandemic, science's disunity is more than ever visible. It is herein asserted that the untimely glimpse into science's inner workings, clashing with the dominant view in popular science, promotes a distorted image of science and hinders people's trust in science. Finally, this article provides wide-ranging recommendations in order to tackle scientism and promote a balanced outlook on science in the fodder consumed by the masses.

Keywords: Scientism, COVID-19 Pandemic, Disunity and Unity of Science, Popularization.

1. Introduction

Science is often at the center of media coverage and public debate. In the last year and a half, it has received greater attention than usual due to the global health crisis provoked by the COVID-19 outbreak. Rightly or not, people and politicians alike now expect scientific knowledge to somehow guide them as they face an unprecedented health crisis in their lifetime. Indeed, there is ample evidence that knowledge, attitudes, and practices about the virus play a major role in fostering adherence to positive behaviors and control measures (Zhong et al. 2020; Hager et al. 2020).

Argumenta 7, 1 (2021): 179-194
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.gaj

The spotlight shone on science brought the issue of dissemination and popularization of scientific outcomes to the fore. These issues bring complex challenges to be tackled, especially in an extraordinary context as the one humanity continues to face twenty months on.

To begin with, the global emergency created an overabundance of information called 'infodemic' (Cinelli et al. 2020), in which it is difficult for people to find trustworthy sources of knowledge to make informed decisions (Porat et al. 2020). In fact, this overexposure to information often heightens people's distress (Holmes et al. 2020) and is positively associated with mental health problems (Porat et al. 2020).

On the practical side, it must be acknowledged that the spreading of scientific knowledge among the public hardly ensures people's adherence to correct behavior: communication does not always achieve its intended outcome. Dagnall and colleagues (2020: 2) pointed out that non-adherence to COVID-19 measures sometimes represents deliberate disregard due to reduced social identity via the creation of an 'us' and 'them.' Social identity can be defined as the sense of self connected to the perceived belonging to a social group: those groups which underestimate, minimize or reject specific information or control measures would inform coherent behaviors in those who self-identify in them. For example, Green and colleagues (2020: 4) noted that, in the early phase of the crisis, the divergent cues sent by U.S. Congressional Democrats and Republicans corresponded with a partisan divide, "with self-identified Democrats reporting significantly more behavioral change than independents and Republicans".

Therefore, it is rather clear that the task of disseminating scientific information is anything but straightforward. On the one hand, dissemination cannot be but a mainstay in institutional recovery plans aimed at facing the pandemic, in that it is a key means to promote paramount attitudes to address the crisis, such as a sense of trust in institutions, a sense of autonomy in decision-making, solidarity, a clear awareness of the limits of what is known and what is not and a sense of social cohesion (Dagnall et al. 2020; Falcone et al. 2020; Porat et al. 2020). On the other hand, scientists and those who popularize science face the problem of meeting the challenge of effectively conveying available scientific outcomes, confronting science's peculiar dynamics and methodological plurality.

The first part of this article deals with the description of the main features of the so-called dominant view in popular science. It is herein claimed that this view is implicitly connected to scientism, an attitude identifying science as the only reliable source of legitimate knowledge. Scientism's philosophical roots are argued to lie at naturalism and at a trivialized neopositivist conception of science. Accordingly, the dominant view is inclined to disregard science's inherent pluralism and cultivates an idealized image of pure and coherent science.

However, in the context of the current pandemic, science's disunity is visible more than ever, both within science (i.e., horizontal disagreement) and in its interface with the public (i.e., vertical misalignment). It is herein argued that the clash between the dominant view and the untimely inside look at science's inner workings promote a distorted image of science.

In the conclusive part, some wide-ranging recommendations are provided, in order to tackle scientism and to promote a fairer attitude toward the role of science in facing the pandemic.

2. The Dominant View in Popular Science

The general aim of popular science is to communicate scientific theories or results to an audience composed of non-specialists. It is worth noting that the term 'non-specialists' does not only refer to a general public or to the so-called non-scientists, rather, the term refers to a large degree of expertise and aims to stress that scientists in a certain field—sometimes even in the same general field—can approach other areas of science as amateurs (Perrault 2013: xiii). This illustrates how thorny is to popularize scientific theories and results. Those who do so (not necessarily scientists but often specialized journalists) must be not only specialists in a specific topic, they must also be able to communicate about it simply but not trivially.

Although the history of popular science is quite recent, starting around the 1600s (Perrault 2013: 37-47), it has undergone changes, especially due to the development of new media and the relationship between scientists and journalists (Dudo 2015: 761-63 and 766). Owing to these changes, at least since the early 1990s, sociologists and communication scientists "have begun to recognize *the fluidity and continuous nature of science communication* rather than its segregate, compartmental division between specialist and popular domains" (Bucchi 2017: 891). In recent times, public audiences appear more engaged in science communication than before (Bucchi 2013: 905). Nonetheless, this engagement "still seems to be lacking among most research institutions in Europe". It can even be argued that much of today's popular science continues to adhere to the so-called dominant view rather than to a model based on public engagement. Hilgartner (Hilgartner 1990: 519; See also Grundmann and Cavaillé 2000: 356) defines this view as a model of communicating science that is based upon an "idealized notion of pure, genuine scientific knowledge against which popularized knowledge is contrasted". The dominant view is depicted as a two-stage model: it begins with scientists developing a genuine, uncontaminated and objective knowledge, a knowledge that represents "the epistemic 'gold standard'" (520); and it ends with popularizers disseminating simplified accounts to the public. At play here is the assumption of the existence of a neat and vast divide between science and 'the rest of the world' (see also Dudo 2015: 764). This means that the rest of the world requires a translation for comprehending scientific knowledge. According to Myers (2003: 266), five claims characterize the dominant view:

1. Scientists and the institutions they pertain to are the authorities on every aspect that constitutes science.
2. Public audiences are assumed to be ignorant on topics with which scientists deal.
3. Scientific knowledge is postulated to go only from science to society, not *vice versa*.[1]
4. Scientific knowledge consists of information contained in a certain number of written statements.
5. When such written information is translated from science to society and public audiences, it must be simplified.

Perhaps the most important issue at stake is the concept of simplification. Clearly popular science must provide simplified accounts to people who do not have any expertise in the field considered. The controversial point is the degree

[1] This is because scientific knowledge is conceived as pure and thus cannot be contaminated by society or other forms of knowledge.

to which this simplification ought to occur. If it is *a priori* assumed that there is a vast gulf between specialists and lay readers, that is to say, "a situation characterized by a hierarchical divide between science and nonscience with technical experts holding the only epistemically valid coin in the realms" (Broks 2006: 46), then this can heavily influence the process of simplification itself. In fact, the aim of this process is to provide simplified accounts of the topics considered, that is, accounts without technicalities. Now, according to the dominant view, because lay readers cannot grasp such technicalities, they have no means to critically discuss the fundamental tenets of the simplified accounts. In other words, for the dominant view such accounts represent the simplified exposition of complex scientific claims that can be only discussed by the experts of the field because non-specialists do not possess an adequate knowledge for understanding their basic assumptions, in that "their claims have not been subjected to the kind of challenges that claims undergo in scientific discourse" (Myers 2003: 269). In this sense, according to the dominant view, popular science is conceived as a mere public relations activity from scientists to lay readers that does not permit any interference or engagement from those who are not specialists (Perrault 2013: 3-6).

In summary, the dominant view conceives the popularization of science as a top-down process through which scientific authorities dispense simplified knowledge without any participation from the public. A clear example can be found throughout the 1985 Royal Society of London report on the public understanding of science (see also Myers 2003: 266; Wynne 1995: 362). In the summary, it is reported that

> Scientists must learn to communicate with the public, be willing to do so, and indeed consider it their duty to do so. All scientists need, therefore, to learn about the media and their constraints and learn how to explain science simply, without jargon and without being condescending (Royal Society of London, 1985: 6).

Scientists must do their best to communicate with public audiences, who are not expected to intervene, criticize or wade into scientific debate. The dominant view of popular science assumes that science lives a life of its own, with no direct contact with other human and public activities (Grundmann and Cavaillé 2000: 353). It conceives of science as existing in a vacuum (379): science cannot be questioned from the outside, only from the inside.

3. Scientism

Throughout the years, sociological and communication science literature has widely discussed the dominant view of popularization and focused on various related issues. Among these, there is the plausibility of the divide between specialists and non-specialists and the idea of scientific knowledge as pure and uncontaminated (Broks 2006); the effects of the power given to scientific authorities over the public and thus the place of science in a democratic society (Perrault 2013); the boundaries between simplification and distortion of the scientific information (Myers 2003); and so on. Sociological and communication science literature generally deals with these topics by discussing detailed historical cases of popular science or focusing on the analysis of surveys or the ways popular science communicates.

Sociological and communication science literature does not appear to be interested in pinning down the philosophical foundations of the dominant view, remaining at a descriptive level useful as a basis for criticism and reflection. However, the assessment of these foundations is not irrelevant because, in their work, popularizers implicitly adopt positions that appear to be scientific, but that are actually philosophical in character. Bunge (2017: 144) argues that thinking there is no connection between science and philosophy is nothing but a myth: it is not plausible to assume "that scientists start from observations, or from hypotheses, and handle them without any philosophical preconceptions". Bunge makes this point by simply offering some examples from the history of science (143-46). This myth refers to the ideal of purity and absence of contamination characterizing the dominant view discussed above. This is a crucial philosophical point because it calls into question the problem of the boundaries of science. As Stenmark argues (2018: 57), the issue is whether any genuine knowledge different from the scientific one actually exists or whether science provides the only reliable manner of obtaining knowledge. If one answers affirmatively, he is a supporter of scientism.

But what is precisely scientism? Very generally, philosophers appear to be divided in those who consider it a thesis or a doctrine (for example, Peels 2017 and 2018; Stenmark 2018) and those who consider it a stance toward science (for example, Bunge 1986 and 2017; Haack 2007; Ladyman 2018). Although this goes far beyond the aims of this paper, it is here considered as a stance, with scientism depicted mainly as an epistemological and methodological point of view (that is, a stance regarding knowledge and the way it is reached) rather than an ontological one (that is, a stance regarding what does and does not exist). This is for two reasons: first, as will be better considered further on, philosophical discussion about scientism occurs mainly at the level of scientific method or the ways through which scientific knowledge is obtained rather than at the level of the questions about the existence of scientific entities; second, apart from some notable exceptions (such as Peels 2017 and 2018 and Stenmark 2018), philosophical literature tends to treat scientism as a stance typical of some supporters of naturalism, no matter here if interpreted at the ontological, epistemological or methodological level. Supporters of scientism tend to be also supporters of naturalism (but not necessarily *vice versa*).

The word scientism is often used in a pejorative sense, not only in philosophical debates but also in public discussions. For example, Haack (2007: 17-18; see also Sorell 1991: 1) depicts scientism as "an exaggerated kind of deference towards science, an excessive readiness to accept as authoritative any claim made by the sciences, and to dismiss every kind of criticism of science or its practitioners as anti-scientific prejudice".[2] Although such a pejorative characterization is not shared by all the critics of scientism (Peels 2018: 30; Stenmark 2018: 59), Haack's definition focuses on a fundamental feature of scientism: the attribution of an authoritative position to science and scientists. As for the dominant view, science provides the most (if not the only) reliable kind of knowledge. It is important to

[2] It is worth noting that Haack (2007: 18) classifies scientism as one of two kinds of confusion we can fall into when we deal with science (and also with popular science, of course). She defines the other 'anti-science', as "an exaggerated kind of suspicion of science, an excessive readiness to see the interests of the powerful at work in every scientific claim, and to accept every kind of criticism of science or its practitioners as undermining its pretensions to tell us how the world is". The problem with each is that they are excessive—scientism in terms of deference, anti-science in terms of suspicion.

stress that here ‘science’ is fully identified with natural science: reliable knowledge cannot come from sources such as common sense, memory, introspection, intuition, religion or, considering only the area of the academic disciplines, the humanities (Peels 2017: 2; Sorell 1991: 9). This is because scientism assumes that science, unlike the other sources of knowledge, has a way of reaching knowledge that is much more advanced and reliable than that of the others: the scientific method or, more precisely, an idealized version of the scientific method typical of the natural sciences (in particular, physics). As Bunge (1986: 25) clearly states, “the scientific method, rather than any special results of scientific research” is “the very kernel of scientism”. For scientism, it is the application of the scientific method that guarantees to obtain the most reliable knowledge. From this claim, it usually follows that, in principle, there are no areas of inquiry that cannot be studied through the methods of natural science. As Ladyman (2018: 113) points out, “everything real can in principle be investigated by scientific methods and no limits should be placed on what science can study”. As a consequence, this leads some supporters of scientism to propose that the scientific method be applied to every sort of human question.

Here, the challenging task of scientism is to define both the scientific method and how it can be potentially applied to every object and field of inquiry, not only to scientific realms but also other areas such as the humanities. If we assume that many cases in the history of science demonstrate that various scientific disciplines use several variations of the scientific method (that is, different methods for different objects), then it is meaningless to think about a single scientific method and, as a consequence, scientism. Thus, in order to adequately defend scientism, we must defend something similar to the old neopositivistic ideal of the unity of method for all the sciences or, at least, the ideal that all science has a well-defined and limited set of methods to be applied. In fact, this ideal is anchored to some basic methodological precepts constituting the backbone of the neopositivist project of science’s unification and representing the ‘hidden bearings’ of those supporting scientism. First, the methods characterizing natural sciences stand as the gold standard of all sciences, underscoring the supposed methodological superiority of a certain family of sciences. Second, scientific disciplines are expected to use strictly empirical procedures in order to postulate general and universal principles. Third, notwithstanding technical differences in investigational methods, scientific statements are supposed to be justified in the same way: deriving from them empirical implications that can be checked intersubjectively. This means that for every private fact there should be a public counterpart which is at the basis of intersubjective agreement, and this is the fourth point (Hempel 1942; see also Gaj 2016). However, certainly scientism cannot be simply assimilated to neopositivism *tout court* (see Sorell 1991: 1-23). Indeed, the latter is a (more or less) consistent philosophical movement, while the former is a general stance about the reliability of knowledge. Here we see scientism takes root in a trivialized version of neopositivism, implicitly endorsing the core message of its main thesis about the unity of science.

In summary, scientism manifests an extreme confidence in knowledge obtained by empirical research: the most reliable knowledge can be produced only through scientific methods, understood as the methods used by natural sciences. This leads to the ideal of unified science, whose feasibility is established under the aegis of one method fitting for all the disciplines appropriately named sciences. From this perspective, all meaningful questions (and answers) about the world must be formulated according to the register of science so understood. Hence, scientism cannot but inherently mingle with naturalism. Indeed, naturalism represents scientism's metaphysical commitment to a world whose features are assumed to be entirely reducible to the categories of the natural sciences.

A key topic related to the core of scientism and the dominant view is the controversial issue of the unity or disunity of the scientific endeavor.

4. The Supposed Unity of Science amid the COVID-19 Pandemic

The connection between the dominant view of popular science and the scientistic stance appears to be quite straightforward. In fact, those popularizers who adopt the dominant view basically depict science as an authority bringing a kind of knowledge that cannot be fully understood and criticized by public audiences. It is crucial to stress that many popular science products typically adopt scientism implicitly: it is difficult to find straightforward formulations of it. A risk here is that popularizers can "come up with woefully inadequate characterizations of key concepts and offer very crude arguments for and against positions that they're discussing" (de Ridder 2014: 23). According to this paper, popularizers tend to implicitly endorse an inadequate characterization of the scientific enterprise as something unified in the name of the scientific method. 'Inadequacy' does not mean that popular science characterizes science as immune to errors or uncontroversial in some stages of its progress. Rather, it means that, in spite of their errors and controversies, scientists who attempt to solve a problem share a common method that allows them to find a common solution or result in the long run. Simply put, the application of a shared and common scientific method allows scientists to agree.

Consider two useful and well-documented sources of information about the COVID-19 pandemic, two recent bestsellers: Richard Horton's *The Covid-19 Catastrophe: What's Gone Wrong and How to Stop It Happening Again* (2020) and Debora MacKenzie's *COVID-19: The Pandemic that Never Should Have Happened, and How to Stop the Next One* (2020). Each author characterizes science in scientistic terms. MacKenzie does this quite explicitly, for example in the following (2020: X):

> What is especially sad for a science journalist like me who writes about disease for a living is that this pandemic has not exactly been a surprise. Scientists have been warning for decades, with mounting urgency, that this was going to happen. And journalists like me have been relaying their warnings that a pandemic is coming and that we aren't prepared.

This excerpt is certainly not wrong: it is plainly true that for a long time scientists were warning us of the possibility of a pandemic scenario similar to the present one (see, for example, Perrin and McCabe 2009). Yet these sentences are implicitly scientistic because they suggest that the scientific community in its entirety and without any controversy shares the same position about the present pandemic and perhaps even about the measures we must take to confront it. The paper will later show that this is a simplistic and unsatisfying narrative about the COVID-19 pandemic.

Although less directly than in MacKenzie's book, examples of scientism are not difficult to find in Horton's book. Both books laud scientific work and, in general, the idea of scientists making a joint effort to deal with the pandemic, without great divisions. Both books provide a historical reconstruction of how scientists discovered SARS-COV-2 and its expressions as a disease and proposed different hypotheses in order to explain it. So, too, both depict the scientific community as a sort of entity using more or less the same methods of research and working in the same direction. This united community seems isolated from 'the rest of the world,' except when things start to go wrong! But what, exactly, goes wrong? The following excerpt from Horton's book puts this question in this manner:

> The global scientific community made an unrivalled contribution to establishing a reliable foundation of knowledge to guide the response to the SARS-CoV-2 pandemic. And yet the management of COVID-19 represented, in many countries, the greatest science policy failure for a generation. What went wrong? (Horton 2020: 41).

On this point, MacKenzie's book is more direct:

> The only real surprise when Covid-19 finally hit was the sheer extent to which most governments simply had not listened to the warnings. We were unable as a planet to muster our considerable scientific understanding of disease in time to soften the blow, never mind preventing it in the first place. And, as I will explain in the coming pages, we could have—at least a lot more than we did. Science didn't actually fail us. The ability of governments to act on it, together, did. Experts had warned about the lack of preparation in addition to the risk of a pandemic itself (MacKenzie 2020: xiii).

Both books place the responsibility for the problems and errors of the management of the pandemic firmly and only on the political sphere: things started to go wrong the moment politicians had to take decisions and create policies. Horton talks about "the gap between the accumulating evidence of scientists and the practice of governments" (Horton 2020: ix). Both he and MacKenzie posit that the scientific side cannot be blamed for anything: scientists were united in their work and proposed the best solutions, but politicians didn't listen to them, for various reasons, ranging from bad intentions to lack of expertise. The image of science is that of an uncontaminated entity corrupted by politicians. The question is whether it is actually all so uncontroversial and clear on the scientific side.

5. Science's Display of Disunity

Well before the declaration of Public Health Emergency of International Concern (January 30, 2020, WHO), scientists throughout the world began jointly to study the behavior of the newcomer in the coronavirus family. As in other cases addressing new phenomena, from the very onset of the pandemic a core set of scientists rapidly formed in order to address issues related to COVID-19: on one side, bottom-up-wise, selected scientists emerged from the scientific community and progressively became deeply involved in research on the virus; on the other side, top-down-wise, policymakers were to designate reliable scholars to collaborate in crafting COVID-19-relevant public policies. The upshot of this essentially social process is the constitution of a group of core-scientists (Collins and Evans 2002) debating on issues related to the many scientific challenges posed by COVID-19. At the moment, they form a community of collectively considered experts who are expected to explore the knowledge frontiers about the virus. Of course, due to the newness of the phenomena under scrutiny, their debates have been characterized by a plurality of voices and, often, disagreement and controversy. For example, key scientists and public-health agencies were late to acknowledge the aerosol transmission of the virus and the consequent benefits of using masks (see Tufekci 2021). Currently, a debate about vaccine safety and transmission from vaccinated individuals continues to take place.

In order to understand how the public could easily understand science's inherent plurality of voices as disunity and fragmentation, it is worth taking into account the dynamics ruling the connections between science and the wider community. Somewhat simplifying their proposal, Collins and Evans (2002) argue that this relation follows approximately this path:

1. Core-scientists debate, confronting different positions about a new subject matter, often via different methods. This involves a high level of uncertainty and diversity of scientific conclusions.
2. As time goes by, scientific disputes normally tend to reach a point where uncertainty and diversity decrease. This entails an inevitable process of simplification and stabilization, by means of which scientific debates are seemingly settled in the eyes of laypeople. "Distance lends enchantment" (ibidem: 246): the more one looks science from a distance, the more unanimous it seems. 'Settled positions' are reached when scientific outcomes are somehow popularized, appropriated in a simplified and coherent manner by the wider scientific community, non-specialists, policymakers and laypeople. This entails a significant reduction of the initial uncertainty characterizing core-scientists' controversies.
3. Despite this outside perception, core-scientists linger long after the wider community believes matters have been settled: science is always open and revisable, at the level of those deeply involved in a specific issue, and is characterized by high levels of uncertainty and plurality of positions and methods of inquiry. High degrees of coherence and certainty characterize popularized versions of science, whereas continuous disputes and plurality inherently characterize any scientific enterprise.

This ideal path does not exactly describe the case for the COVID-19 pandemic. Being a new and demanding challenge, science cannot but fail to rapidly provide relatively stable outcomes for the wider community. As noted above, time is required before core-scientists' conclusions are stabilized as a result of exposure to the wider community. On the contrary, the conflicting dynamics of science have never been as visible to laypeople and the media as they have been during this pandemic, at a time when solid scientific answers are more yearned for than ever. The problem here is not that laypeople and policymakers may not be adequately prepared to understand scientific results, as the dominant view would suggest; rather, they may not be adequately prepared to confidently address science's characteristic plurality of voices in approaching a new phenomenon.

Indeed, the untimely exposure to scientific inner workings may prejudice people's attitudes toward science and bring potential damage to its credibility. While experts may consider disagreements to be part of the scientific process, laypeople may have a different perception. The uncertainty, fluidity and disunity characterizing core-scientists' controversies may violently clash with the expectations of both policymakers and laypeople, who generally hope for unanimous and definite outcomes from science. Real-life decisions are mostly formulated as binary choices (yes/no, do that/do not do that): so, it is likely that science is considered as a useful tool to the extent it substantially contributes to disentangling such knots (Collins and Evans 2002; Nichols 2017).

In this scenario, science may appear far less united and robust than expected and may be no longer viewed as a source of confidence (Collins and Evans 2002): uncertainty and distrust may spread in the wider community, fostering contentious attitudes toward science (Kosolosky and Van Bouwel 2014). Indeed, even in non-emergencies, laypeople tend to use overly narrow attributions to make sense of scientific disagreements, overlooking the irreducible uncertainty of the world as a relevant source of scientific dispute. Even though education and available cognitive resources play a role, people tend to favor inferences according to which uncertainty stems from either incompetence or bias on the part of the experts (Dieckmann et al. 2017). Such effects might well worsen in a situation where the public is prematurely exposed to early debates among core-scientists and the uncertainty level is high.

So, how does science's disunity manifest in the context of the current pandemic? This article argues that horizontal disagreement (HD) is to be distinguished from vertical misalignment (VM).

HD is defined as the expected discordance among core-scientists when addressing a relatively new topic or problem. Various theoretical and methodological positions compete. The accuracy of scientific outcomes (that is, the preciseness of the answer given to a certain query) is at stake here (Kosolosky and Van Bouwel 2014). In the early phases of investigation of a relatively new phenomenon, core-scientists struggle to reach the most thorough understanding possible. However, it is likely that scientific accuracy is partial and rapidly growing in these phases, possibly conveying the idea that science is unstable and ever-changing. The notion of robustness, or lack thereof, might account for this. Robustness is the idea that hypotheses about entities and processes are better supported if they are detectable, derivable, producible in a variety of independent ways, i.e., via multiple techniques and methods relying on different background assumptions (Eronen 2015; Miller 2013; Stegenga 2009). A scientific conclusion deserves to be defined as robust when many independent threads of evidence progressively converge in an intelligible form and a picture gradually emerges and becomes detectable. Before this threshold, those independent threads of evidence may appear as fragmented, if not opposing, positions within the scientific arena. So, when a line of research lacks robustness due to its immaturity, what is seen is fragmentation. In the present situation, laypeople are often exposed to HD well before science's different voices may converge in a (more or less) coherent picture. This might easily create an appearance of incoherence and disunity. In the context of the present pandemic, some instances of HD have been, and are, clearly visible to the wider community. These include: disputes about whether the origins of COVID-19 were artificial vs. natural (Chaturvedi, Ramalingam, Singh 2020; Andersen, Rambaut, Lipkin, Holmes, Garry 2020); and the mystery of long COVID, a varied syndrome that can have long-term disabling effects, about which very little is known (Sollini et al. 2021).

VM (vertical misalignment) regards the relationship between scientists and the public. As already noted, evident disagreement among scientists may be easily taken as a sign of incompetence or bias influence (Dieckmann et al. 2017), favoring mistrust and disbelief (Collins and Evans 2002; Nichols 2017): "everyone gets to see the soft flesh of the scientific fruit and the familiar passions and arguments that constitute it" (Collins and Evans 2002: 248). This markedly influences the credibility of visibly ever-changing scientific conclusions. When it comes to the interface between core-scientists and the community at large, VM deals with the notion of adequacy. Adequacy points at what an explainee expects of a scientific answer and, thus, deals with how scientific outputs and the explainee's epistemic interests reciprocally fit. Adequacy concerns the congruity of scientific outcomes with the interests or *desiderata* at play (Kosolosky and Van Bouwel 2014). Borrowing an example from the literature on maps, the need to orient people in a specific environment, e.g., a subway system, requires maps with an *adequate* degree of accuracy, precisely based on the users' epistemic needs. The adequacy of a map depends on its inclusion/exclusion of specific features based on the purpose for which the map is being made (Giere 2006): not all maps (supposedly having different degrees of accuracy) would fit the users' needs. More precisely, VM deals with the relation between an incomplete and still growing accuracy—as a characterizing feature of HD—and adequacy. In fact, adequacy can be achieved to the extent that a certain degree of accuracy has been reached among core-scientists. In the present situation, the ever-growing accuracy and the instability of scientific conclusions characterizing HD make adequacy a hard goal to reach, fostering a state of VM where people may be easily disoriented by science's provisional answers. Of the several instances of VM emerging from the current pandemic, two stand out for their ability to highlight the misalignment between science and the public.

One regards the early use of masks. Late to ascertain the usefulness of wearing masks, even considering that imposing the wearing of masks *per se* poses few downsides (Tufekci 2021), science-wise scientists waited for a high level of accuracy and certain evidence before communicating the effectiveness of this measure. In doing so, the public might have been confused by science's reticence to express a clear opinion in the early phase of the pandemic. Accordingly, relevant recommendations for protecting people from airborne transmission came late and with less credibility than required.

The second case regards different ways to understand and communicate how to measure the contagiousness of the virus. After many months of collecting data, scientists discovered that parameter R0 (an indicator of the average number of secondary infections caused by every infected individual) is an insufficient index for understanding how the virus spreads. Because COVID-19 tends to spread in clusters, the average is not useful for understanding its distribution. On the contrary, parameter *k*, which might be not as familiar as R0, is an adequate measure to account for the behavior of an over-dispersed pathogen such as COVID-19 (Endo, Abbott, Kucharski, and Funk 2020; Hasan et al. 2020; Tufekci 2020). Public exposure to growing scientific accuracy about the behavior of the virus may have puzzled laypeople and had a late and suboptimal influence on the design of appropriate safety measures and on people's compliance.

These are but two examples in which science's provisional outcomes disoriented the public, not providing univocal information and unwittingly disseminating the image of a disunited science.

Both HD and VM convey an image of science as an inherently pluralistic, disunited, ever-progressing endeavor. Indeed, this is exactly science's nature, something quite different—and perhaps even more intriguing and sophisticated—than that proposed by scientism. The clash between the dominant view and the view conveyed by the untimely exposure to science's early controversies may drastically hinder a mature consideration of its virtues and limits.

In the concluding section, the article will suggest some recommendations to promote a balanced image of science, against scientism and its popularized counterpart, the dominant view. This aims to educate the public to science's inherent pluralistic and provisional status, which must be understood as a desirable trademark feature, rather than a sign of unreliability.

6. Concluding Remarks

The main aim of scientism is to promote an image of science as pure and able to provide solutions to every problem through the application of a well-established and recognized method or set of methods. This is plainly wrong: the COVID-19 pandemic continues to show that science is far from perfect, full of controversies and difficulties, with a plurality of methods and points of view. Further, science does not stand apart from the world: indeed, far from existing in a vacuum, science is in a continuous exchange with the public and everyday life—and the present pandemic is perhaps the most dramatic proof of this fact. For these reasons, the article argues that one of the preliminary tasks of popularization must be to demonstrate and explain how science is pluralistic and integrated in social discourse. It must provide people the means to debunk the various myths regarding science and to form a critical opinion about it.

At the operative level, this might mean various different things. To begin with, consider a central methodological procedure commonly used in science: the rejection of the null hypothesis, the way through which knowledge is produced and is gained credibility in academia. Usually, scientists are interested in testing the experimental hypothesis (i.e., the prediction that the manipulation of the independent variable will have some effect on the dependent variable) by rejecting the opposite null hypothesis, namely, that the prediction is wrong and that there is no relationship between variables. The experimental hypothesis can be accepted only if it excludes that the results obtained accordingly occurred by chance (see Field 2005). Although this is a common scientific procedure, laypeople may misunderstand it. First, scientific outcomes may be rhetorically presented as the exclusion of the possibility that one variable has a relationship with another variable, until the relationship is not proven by the (definitive)[3] rejection of the null hypothesis. This would convey the idea that some conclusions are not (this far) 'scientifically proven', even if they are highly probable on the basis of current knowledge. As Tufekci reported,

> On January 14, 2020, the WHO stated that there was 'no clear evidence of human-to-human transmission.' It should have said, 'There is increasing likelihood that human-to-human transmission is taking place, but we haven't yet proven this' (Tufekci 2021).

[3] Here, the adjective 'definitive' is not to be literally understood. Rather, it refers to values which are statistically significant. Normally, a 95 percent probability that the relationship between the variables is genuine vs. a 5 percent probability that the relationship occurs by chance, is considered a sufficient value to argue that scientists face a real effect or that there exists a true relationship between variables (Field 2005).

Accordingly, scientific conclusions may appear as more uncertain that they actually are, if only available information is carefully taken into consideration. Still following Tufekci,

> Later that spring, WHO officials stated that there was 'currently no evidence that people who have recovered from COVID-19 and have antibodies are protected from a second infection,' producing many articles laden with panic and despair. Instead, it should have said: 'We expect the immune system to function against this virus, and to provide some immunity for some period of time, but it is still hard to know specifics because it is so early' (Tufekci 2021).

Popularizers must be aware of these methodological features when they design their dissemination strategies. Indeed, they should raise awareness on science's specific workings, in order to contain scientism and strengthen people's trust in science. Accordingly, they must provide basic epistemological notions to their readers in order to explain how science works (Dieckmann et al. 2017: 335). In particular, science's characteristic disunity, its probabilistic nature and the inherent limitations of scientific knowledge should be made known to laypeople: in familiarizing the public, popularizers ought to convey a clear image of the scientific enterprise as uncertain and provisional (Porat et al. 2020: 9), yet valuable. Moreover, popularizers should always keep in mind the needs of the various audiences and contexts for which the information is intended, in that these factors considerably influence the understanding of information and the adhering to safe behaviors (Bucchi 2017: 890; Kosolosky & Van Bouwel 2014; Porat et al. 2020: 8). Lastly, popularizers should nurture the skill of communicating honestly what is known and what is not, conscientiously indicating the aims and general strategy pursued by scientific research (Porat et al. 2020: 10).[4]

These are only few recommendations for improving popular science. They deserve to be further developed by the joint efforts of scientists, philosophers and popularizers. The main contribution of this article to the topic of popular science is to suggest not to confound science with scientism: scientism offers a distorted image, if not a caricature, of what science really is. Science is disunited, not united—and we should not be afraid of it.

References

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F. 2020, "The proximal Origin of SARS-CoV-2", *Nature Medicine*, 26, 450-52.

Broks, P. 2006, *Understanding Popular Science*, New York: Open University Press.

Bucchi, M. 2013, "Style in Science Communication", *Public Understanding of Science*, 22, 8, 904-15.

[4] As a reviewer stressed, popularizers often exhibit a kind of "epistemic irresponsibility" (Magnani 2017: 161-97. See also Park 2020) by spreading an image of science as a form of static knowledge, that is, something contrary to the so-called knowledge in motion. By underestimating the epistemic and philosophical topics underlying the scientific enterprise, they tend to describe scientific research as not multidisciplinary, interdisciplinary and transdisciplinary (Magnani 2017: 161) and thus unable to foster "good human abductive creative reasoning" (Magnani 2017: 167). Very briefly, popularizers promote an image of science as fake or, at the very best, uninteresting.

Bucchi, M. 2017, "Editorial. Credibility, Expertise and the Challenges of Science Communication 2.0", *Public Understanding of Science*, 26, 8, 890-93.

Bunge, M. 1986, "In Defense of Realism and Scientism", *Annals of Theoretical Psychology*, 4, 23-26.

Bunge, M. 2017, *Doing Science in the Light of Philosophy*, Singapore: World Scientific Publishing.

Chaturvedi, P., Ramalingam, N., and Singh, A. 2020, "Is COVID-19 Man-Made?", *Cancer Research Statistics and Treatment*, 3, 284-86.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F., Scala, A. 2020, "The COVID-19 Social Media Infodemic", *Scientific Reports*, 10, 165-98.

Collins, H.M., Evans, R. 2002, "The Third Wave of Science Studies: Studies of Expertise and Experience", *Social Studies of Science*, 32, 2, 235-96.

Dagnall, N., Drinkwater, K.G., Denovan, A., Walsh, R.S. 2020, "Bridging the Gap between UK Government Strategic Narratives and Public Opinion/Behavior: Lessons from COVID-19", *Frontiers in Communication*, 5, 1-8.

de Ridder, J. 2014, "Science and Scientism in Popular Science Writing", *Social Epistemology Review and Reply Collective*, 3, 129, 23-39.

Dieckmann, N.F., Johnson, B.B., Gregory, R., Mayorga, M., Han, P.K., and Slovic, P. 2017, "Public Perceptions of Expert Disagreement: Bias and Incompetence or a Complex and Random World?", *Public Understanding of Science*, 26, 39, 325-38.

Dudo, A. 2015, "Scientists, the Media, and the Public Communication of Science", *Sociological Compass*, 9, 9, 761-75.

Editorial 2020, "Meeting the Challenge of Long COVID", *Nature Medicine*, 26, 1803.

Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., Kucharski, A.J., and Funk, C. 2020, "Estimating the Overdispersion in COVID-19 Transmission Using Outbreak Sizes outside China", *Wellcome Open Research*, 5, 67 [version 3; peer review: 2 approved].

Eronen, M.I. 2015, "Robustness and Reality", *Synthese*, 192, 3961-77.

Falcone, R., Colì, E., Felletti, S., Sapienza, A., Castelfranchi, C., and Paglieri, F. 2020, "All We Need Is Trust: COVID-19 Outbreak Reconfigured Trust in Italian Public Institutions", *Frontiers in Psychology*, 11, 1-17.

Field, A. 2015, *Discovering Statistics Using SPSS*, 2nd Edition, London, Thousand Oaks and New Delhi: Sage.

Gaj, N. 2016, *Unity and Fragmentation in Psychology: The Philosophical and Methodological Roots of the Discipline*, London and New York: Routledge.

Giere, R.N. 2006, *Scientific Perspectivism*, Chicago and London: The University of Chicago Press.

Green, J., Edgerton, J., Naftel, D., Shoub, K., Cranmer S.J. 2020, "Elusive Consensus: Polarization in Elite Communication on the COVID-19 Pandemic", *Science Advance* 6, 1-5.

Grundmann, R., Cavaillé, J.P. 2000, "Simplicity in Science and its Publics", *Science as Culture* 9, 353-89.

Haack, S. 2009, *Evidence and Inquiry. A Pragmatist Reconstruction of Epistemology*, 2nd Edition, Amherst: Prometheus Books.

Hager, E., Odetokun, I.A., Bolarinwa, O., Zainab, A., Okechukwu, O., Al-Mustapha, A.I. 2020, “Knowledge, Attitude, and Perceptions towards the 2019 Coronavirus Pandemic: A Bi-National Survey in Africa”, *PLOS ONE*, 1-13.

Hasan, A., Susanto, H., Kasim, M.F. et al. 2020, “Superspreading in Early Transmissions of COVID-19 in Indonesia”, *Scientific Report* 10, 223-86.

Hempel, C.G. 1942, “The Function of General Laws in History”, *Journal of Philosophy*, 39, 35-48.

Hilgartner, S. 1990, “The Dominant View of Popularization: Conceptual Problems, Political Uses”, *Social Studies of Science*, 20, 519-39.

Holmes, E.A., O’Connor, R.C., Perry, V.H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Silver Cohen, R., Everall, I., Ford, T., John, A., Kabir, T., King, K., Madan, I., Michie, S., Przybylski, A.K., Shafran, R., Sweeney, A., Worthman, C.M., Yardley, L., Cowan, K., Cope, C., Hotopf, M., Bullmore, E. 2020, “Multidisciplinary Research Priorities for the COVID-19 Pandemic: A Call for Action for Mental Health Science”, *Lancet Psychiatry*, 7, 547-60.

Horton, R. 2020, *The COVID-19 Catastrophe: What’s Gone Wrong and How to Stop It Happening Again*, Cambridge: Polity Press.

Kosolosky, L., Van Bouwel, J. 2014, “Explicating Ways of Consensus-Making: Distinguishing the Academic, the Interface and Meta-Consensus”, in Martini, C. and Boumans, M. (eds.), *Experts and Consensus in Social Sciences*, Dordrecht: Springer, 71-92.

Ladyman, J. 2018, “Scientism with a Humane Face”, in de Ridder J., Peels, R., and van Woudenberg, R. (eds.), *Scientism: Prospects and Problems*, Oxford: Oxford University Press, 106-26.

Mackenzie, D. 2020, *COVID-19. The Pandemic that Never Should Have Happened and How to Stop the Next One*, Paris: Hachette Books.

Magnani, L. 2017, *The Abductive Structure of Scientific Creativity: An Essay on the Ecology of Cognition*, New York: Springer.

Miller, B. 2013, “When is Consensus Knowledge Based? Distinguishing Shared Knowledge from Mere Agreement”, *Synthese*, 190, 1293-1316.

Myers, G. 2003, “Discourse Studies of Scientific Popularization: Questioning the Boundaries”, *Discourse Studies*, 5, 2, 265-79.

Nichols, T. 2017, *The Death of Expertise: The Campaign against Established Knowledge and Why it Matters*, New York, NY: Oxford University Press.

Park, W. 2020, Review of Magnani 2017, *Transaction of the Charles S. Pierce Society*, 56, 3, 456-65.

Peels, R. 2017, “Ten Reasons to Embrace Scientism”, *Studies in History and Philosophy of Science*, Part A 63, 11-21.

Peels, R. 2018, “A Conceptual Map of Scientism”, in de Ridder J., Peels, R., and van Woudenberg, R., (eds.), *Scientism: Prospects and Problems*, Oxford: Oxford University Press, 28-56.

Perrault, S. 2013, *Communicating Popular Science: From Deficit to Democracy*, London: Palgrave Macmillan.

Porat, T., Nyrup, R., Calvo, R.A., Paudyal, P., Ford, E. 2020, “Public Health and Risk Communication during COVID-19: Enhancing Psychological Needs to Promote Sustainable Behavior Change”, *Frontiers in Public Health*, 8, 1-15.

Royal Society of London (report of) 1985, *The Public Understanding of Science*, London: The Royal Society.

Sollini, M., Morbelli, S., Ciccarelli, M., Cecconi, M., Aghemo, A., Morelli, P., Chiola, S., Gelardi, F., and Chiti, A. 2021, "Long COVID Hallmarks on [18F]FDG-PET/CT: A Case-Control Study", *European Journal of Nuclear Medicine and Molecular Imaging*, 48, 10, 3187-97.

Sorell, T. 1991, *Scientism: Philosophy and the Infatuation with Science*, London: Routledge.

Stegenga, J. 2009, "Robustness, Discordance, and Relevance", *Philosophy of Science*, 76, 5, 650-61.

Stenmark, M. 2018, "Scientism and Its Rivals", in de Ridder J., Peels, R., and van Woudenberg, R. (eds.), *Scientism: Prospects and Problems*, Oxford: Oxford University Press, 57-82.

Tufekci, Z. 2020, "This Overlooked Variable is the Key to the Pandemic: It's not R", *The Atlantic*, https://www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/.

Tufekci, Z. 2021, "5 Pandemic Mistakes We Keep Repeating: We Can Learn from Our Failures", *The Atlantic*, https://www.theatlantic.com/ideas/archive/2021/02/how-public-health-messaging-backfired/618147/.

Wynne, B. 1995, "Public Understanding of Science", in Jasanoff S., G.E. Markle, J.C. Petersen, and Pinch, T. (eds.), *Handbook of Science and Technology Studies,* Thousand Oaks: Sage, 361-88.

Zhong, B.L., Luo, W., Li, H.M., Zhang, Q.Q., Liu, X.G., Li, W.T, and Li, Y. 2020, "Knowledge, Attitudes, and Practices towards COVID-19 among Chinese Residents during the Rapid Rise Period of the COVID-19 Outbreak: A Quick Online Cross-Sectional Survey", *International Journal of Biological Sciences*, 16, 1745-52.

Online References

https://www.who.int/news-room/spotlight/a-year-without-precedent-who-s-covid-19-response (retrieved February 23, 2021).

Keeping Doors Open:
Another Reason to Be Skeptical of Fine-Based Vaccine Policies

Stefano Calboli and Vincenzo Fano

University of Urbino Carlo Bo

Abstract

An impressive effort by the scientific community has quickly made available SARS-CoV-2 vaccines, indispensable allies in the fight against COVID-19. Nevertheless, in liberal democracies, getting vaccinated is an individual choice and a not-negligible number of persons might turn out to be vaccine refusers. Behavioral and Cognitive (B&C) scientists have cast light on the key behavior drivers of the vaccine choice and suggested choice architectures to boost vaccine uptake. In this paper, we identify a somehow neglected psychological phenomenon, that it is reasonable to believe to hamper the vaccine uptake whereby fine-based coercive policies are in place. We begin by presenting the default effect, peer pressure, and the case versus base-rate effect as examples of psychological mechanisms relevant for vaccine choice. We show interventions on the choice environment conceived to manipulate such mechanisms (§1). Next, we focus on what B&C scientists have investigated as well the conditions under which monetary disincentives become ineffective policy measures. To do this, we discuss in detail the case of the crowding-out effect (§2). In section 3 we present the original point of the paper. We argue that imposing monetary disincentives on vaccine hesitant could turn out to be ineffective also because of the human tendency to keep options open, albeit doing so bears some cost. In section 4 we draw an experiment aimed to begin testing whether the tendency to keep options open factually plays a role within the context of the vaccine choice (§4). Finally, concerning the COVID-19 emergency, we defend an attitude of epistemic humility in translating behavioral and cognitive research results into policy suggestions (§5).

Keywords: Keeping doors open, Behavioral and cognitive sciences, Evidence-based policy, COVID-19 emergency, Vaccine hesitancy.

1. Introduction

The body of knowledge made available by behavioral and cognitive (henceforth B&C) scientists is of utter importance for those policy-makers whose aim is to promote vaccine uptake. This is because in modern liberal democracies institutions cannot physically constrain competent citizens to get vaccine jabs since it

Argumenta 7, 1 (2021): 195-214
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.cal

would violate the principle by which a medical procedure can be undertaken only if consent is given. So, both being vaccinated and refusing vaccines are options on which citizens choose freely, even if coercive policies are in place. Indeed, when scholars discuss the implementation of mandatory vaccination policies in liberal democracies, they refer to policies whereby incentives and disincentives of different nature and severity are adopted and never to cases in which citizens are compelled by force. Coercive measures may correspond to financial incentives and disincentives. For instance, in 2015, the Australian government introduced the so-called "No Jab No Pay" policy, for which financial child support was withheld from parents of unvaccinated children when medical exemptions were excluded (Trentini et al. 2021).[1] Furthermore, coercive policies can be based on non-financial incentives and disincentives, as in the case of the so-called "Lorenzini Law". In 2017, in response to alarmingly decreasing vaccine coverage rates and to several measles outbreaks (Siani 2019), the Italian government approved the 119/2017 law, aka the "Lorenzini Law". The Lorenzini Law made ten vaccines mandatory for children through requiring proof of vaccination as a condition to see children admitted into preschools, day-care centers, or primary schools (Signorelli et al. 2018). In addition, the 119/2017 law included a further, and arguably more severe, coercive tool: administrative sanctions which were imposed on the families of unvaccinated children. However, it should be noted that these sanctions have been meted out very occasionally (Magnano 2019). Finally, over the course of the years, governments around the world have exploited an even more coercive policy instrument to curb the vaccine refusal rate, namely the incarceration of vaccine decliners (Gravagna et al. 2020).[2]

Inasmuch the vaccine uptake by adults is a free choice, although more or less effectively influenced by policies, the role that B&C sciences could play in shaping vaccine uptake is twofold. B&C sciences shed light on the decision processes underpinning the vaccine choice, revealing which psychological mechanisms ease and which hamper the vaccine uptake. B&C scientists also investigate which modifications of the choice environment influence such mechanisms. Variations of this sort are called "nudges": roughly speaking modifications of the choice environment inhabited by decision-makers that do not imply any form of coercion, ban or significant economic incentives or disincentives (Thaler and Sunstein 2008).[3] Concerning vaccines, B&C scientists have investigated several psychological mechanisms turned out to be relevant. Although it is beside the point of this article to exhaustively list them, we discuss some exam-

[1] It should be noted that the "No Jab No Pay" policy is slightly different from classic financial disincentives. Indeed, this Australian policy does not entail the incurring of a cost, instead preventing access to financial supports. The "No Jab No Pay" set a peculiar choice environment that could bring some advantages in view of the endowment effect (Kahneman et al. 1990).

[2] For overviews on compulsory vaccination policies around the world see also Walkinshaw 2011.

[3] Although some scholars categorically reject the use of nudges, overall the use of such soft interventions is considered legitimate in liberal democracies, being regarded as a measure that preserves individual freedom. However, the conditions under which this is factually the case is debated, see Goodwin 2012, Grüne-Yanoff 2012 and Schmidt 2017.

ples to realize the role nudges play in enhancing vaccine policies.[4] For instance, policy-makers can effectively take advantage of B&C research on both the strength of the default setting and the allure of social norms in order to promote vaccine uptake.

The default effect emerges when a specific option, within a certain set, is more likely to be chosen if it is the default option, viz. the option with which the chooser ends up if nothing is done. The default effect has been found out to be relevant in shaping human decisions in a wide range of contexts (Jachimowicz et al. 2019), for instance, decisions relevant to health (Halpern et al. 2007) such as enrolling in and adhering to a behavioral intervention program for patients with poorly controlled diabetes (Aysola et al. 2016) or donating organs (Johnson and Goldstein 2003). Making an appointment for vaccine uptake is no exception and the role of the default effect in driving the vaccine appointment has been repeatedly confirmed, although it is unclear if it translates into an increase of actual vaccine uptake (Chapman et al. 2010; Lehmann et al. 2016). Hence policy-makers can effectively take advantage of the default effect through opt-out strategies, for which vaccine appointments are set up by default. In this case, no action is required to make the appointment; instead it is cancelling an appointment that requires action.

The strength of social norms and peer pressure in favoring health-promoting behavior is another aspect underlined by B&C research. Our behaviors are conditioned on what we think other people believe it is right to do and on what we think they factually do. Indeed, knowing that the vast majority of our reference social network believe that a certain behavior is right and act accordingly spurs human beings to behave the same way (Abrams et al. 1990; Cialdini and Goldstein 2004; Bicchieri and Dimant 2019; Bicchieri 2016). The impact of social norms on human behaviors has been detected in a wide range of contexts. Some of them are unrelated to the choices relevant to curb the spread of the SARS-CoV-2 virus, as for instance the use of hotel towels (Goldstein et al. 2008) or the decisions taken facing economic games (Fischbacher et al. 2001); others that are deeply related to the current pandemic include wearing medical face masks and respecting physical distancing (Nakayachi et al. 2020; Bicchieri et al. 2020).[5] Norm-nudging precisely leverages social norms and peer pressure. It indeed consists of implementing communication strategies that emphasize either that the vast majority of the reference network is behaving as policy-makers desire or, at least, that an increasing percentage of that network is opting for the target behavior (Sparkman and Walton 2017). A creative and effective kind of norm-nudging consists of providing items that allow citizens who have already chosen the target option to signal their choice to others (Bond et al. 2012; Hayes et al. 2015; Yoeli et al. 2013). Such items, for example the "I Got my COVID-19 Vaccine" Facebook profile frame, has led observers to infer the popularity of the target choice. In general, peer pressure has been confirmed to

[4] For lists of research results from B&C sciences potentially relevant for the COVID-19 emergency, see Bavel et al. 2020, Wood and Schulman 2021, World Health Organization 2020, and finally Lunn et al. 2020.

[5] We prefer “spatial distancing” over the widespread and misleading “social distancing”. The COVID-19 emergency calls for limiting close physical human connections and not for restricting social interactions. Nowadays, we can reach out to other persons through many communication tools that do not entail physical contact (see Abel and McQueen 2020).

be one of the key behavior drivers of the vaccine choice (Bish et al. 2011; Xiao and Borah 2020), and it appears the same applies to the case of the SARS-CoV-2 vaccine uptake (Thaker 2020, Preprint). Hence, communication strategies promoting citizens' perception that their peers are vaccinated (Bruine de Bruin et al. 2019; Felletti 2020) as well strategies that take advantage of social signaling tools, are ways to effectively promote vaccine uptake. For the same reason, communicating social disapproval for a specific behavior increases the chance that an alternative conduct, endorsed by peers, will be preferred. By extension, evidence from research on social norms and peer pressure suggests that communication strategies that lead citizens to infer the popularity of the behavior that policy-makers want to counter are to be avoided (Schultz et al. 2007). In the case of vaccines, this translates into public information campaigns that emphasize the support for vaccine uptake instead of giving voice to the few, yet vocal, naysayers, to discredit them.

B&C research reveals, as well, the psychological mechanisms and the features of the choice environment that keep citizens away from vaccines. In this respect, B&C scientists strongly warn policy-makers about an especially powerful damper for vaccine uptake, namely the case versus base-rate effect, for which human beings irrationally overweight event-specific information and underweight base-rate statistics (Lynch and Ofir 1989). Such an effect seems to be related to the availability heuristic, where the likelihood assigned by human beings to an event heavily depends on the readiness with which instances of that event come to their mind (Tversky and Kahneman 1973). The case versus base-rate effect raises concerns regarding the vaccine uptake rate particularly because mass media, users of social networks and circles of acquaintances are constantly displaying and commenting on both rumors and anecdotes of adverse events that occurred soon after the uptake of SARS-CoV-2 vaccines. As a result, B&C scientists have recommended physicians counteract the detrimental effect of anecdotes about medical issues that emerged after the vaccine uptake by turning the allure of anecdotes to their advantage. For example, they suggest augmenting statistical explanations with anecdotes about the social benefits that result from vaccination uptake (Wood and Schulman 2021). Hence, B&C research enhances our understanding of the drivers of the behaviour relevant in fighting the spread of the SARS-CoV-2 virus. However, research results from B&C sciences not only pave the way toward the use of unconventional policy means such as nudges, but also throw light on the conditions that guarantee the effectiveness of traditional means such as, for instance, monetary disincentives. This is the focus of the following section.

2. Skepticism about Fining Vaccine Refusers

The standard economic theory of punishment makes clear predictions on how fines influence human behaviors. Let's consider the case in which two options are available, *x* and *y*, and one of them, say *x*, involves the payment of a fine. Let's refer initially to the choosers who would have maximized their utility with *x* if any fine was imposed. The theory predicts that the fine could have one of two possible consequences over them. On the one hand, the punishment could be substantial enough to determine a switch of the preferences and lead them to maximize utility choosing *y* instead. On the other hand, the punishment could not be enough substantial to generate a preferences switch, hence they still maximize their utility choosing *x* over *y*, despite the fine. The standard economic theory of punishment further predicts that a fine, regardless of its entity, does not make any difference in the decision processes of those who would have maximized their utility choosing *y* regardless of whether a punishment was imposed on *x* (Becker 1968). B&C scientists, however, point out that the effects of punishment over human behaviors are not so straightforward. There are indeed several cases in which imposing punishments over a certain behavior promotes it instead of discouraging it. In other words, there are conditions under which monetary punishments could easily backfire (Xiao 2018).

Getting inoculated is perceived as a prosocial choice by many, being a choice that preserves the health of others, especially in that it contributes to reaching herd immunity, which is a public good (Buttenheim and Asch 2013). Regarding the specific case of the SARS-CoV-2 vaccines, so far it is uncertain whether those who get the vaccine indirectly protect unvaccinated people or whether they keep spreading the virus to the unvaccinated, even if recent research offers grounds for optimism (Petter et al. 2021; Levine-Tiefenbrun et al. 2021; Mallapaty 2021. However, see also Subbaraman 2021). However, regardless of whether the SARS-CoV-2 vaccines either stop spread transmission of the virus altogether, or reduce spread significantly, or do not reduce it at all, receiving the SARS-CoV-2 vaccine is in any case rightly perceived as a prosocial choice. Firstly, getting the SARS-CoV-2 vaccine is a prosocial choice because the more persons who are vaccinated, the fewer opportunities the SARS-CoV-2 virus has to replicate itself, reducing the emergence of new potentially worrisome variants (McCormick et al. 2021). Secondly, the persons in need of treatment for the Coronavirus disease are causing an atypical and massive flow to hospitals, compromising the function of health systems around the world. The resources of health systems, in terms of medical personnel and beds, are limited. Getting the vaccine is, therefore, a prosocial choice in that it would reduce the number of such patients, preventing the draining of such currently highly demanded resources (Maringe et al. 2020).

The prosocial value of behaviours has been found to motivate human choices and vaccine uptaking seems to be no exception (Betsch et al. 2013). Consequently, communication strategies aimed at emphasizing the social benefits that spring from vaccination prod citizens to get vaccinated (Quadri-Sheriff et al. 2012; Korn et al. 2018). Unfortunately, the intrinsic motivations to behave prosocially are typically crowded out when monetary disincentives are imposed on those who prefer the course of action which does not imply social benefits. Such a phenomenon is an instance of the so-called "motivation crowding-out effect". It has been found to be relevant in several contexts and to be triggered even when monetary incentives reward behaviours motivated by "moral sentiments" (Fehr and Falk 2002; Gneezy et al. 2011; Ariely et al. 2009; Bowles 2008; Mellström and Johannesson 2008; Fehr and Gächter 2002; Fehr and Rockenbach 2003; Gneezy and Rustichini 2000; Calabuig et al. 2016). Vaccine uptake is not immune to the motivation crowding-out effect (Buttenheim and Asch 2013; Madrian 2014) and experts in the B&C sciences have already raised doubts about policies that either introduce monetary disincentives for SARS-CoV-2 vaccine decliners or monetary incentives to award vaccine takers (Santos Rutschman 2020; Schmelz 2020). Instead, they suggest adopting non-financial incentives and disincentives. For instance, Richard Thaler—one of the fathers of behavioral economics—has proposed introducing COVID-19 health passports that would ensure the possibility of gaining access to public spaces as a perk for vaccine takers (Thaler 2021). This measure has already been taken into consideration by several governments (Phelan 2020) and already implemented in some cases (Ministry of Health of Israel, n.d.; European Parliament and European Council 2021).

It could be said that the concerns based on the detrimental motivation crowding-out effect are not really that worrying. Indeed, policy-makers can solve the root of the problem imposing financial disincentives severe enough to make the expected cost of refusing vaccination so high as to lead most citizens to get vaccinated. Nevertheless, to impose fines high enough to overcome the expected gain of refusing the vaccine is inadvisable for two reasons. Firstly, it requires costly monitoring (Bicchieri et al. 2021), and secondly, it could foster the perception of fines as unfair acts, resulting in a hostile atmosphere and eventual retaliation (see Xiao 2018). Hence, what is factually viable for policy-makers is weak fines that, unfortunately, could potentially trigger the motivation crowding-out effect. However, this is not the only effect at play.

Our original contribution to the discussion is to show that, in light of B&C research, it is reasonable to believe that there is another major, and somehow neglected, mechanism that justifies skepticism about imposing weak fines on vaccine hesitant. This mechanism, rather than concerning the crowding-out effect, consists of the human tendency to keep options open, when the chance is given, even if doing so bears some cost.

3. Fines as Opportunities to Keep Options Open

In the present section, we suggest that research results from B&C sciences make it sensible to conjecture that there is a reason other than the crowding-out effect to be doubtful about the effectiveness of using of monetary disincentives to prevent vaccine refusing. Discussion of this reason is made urgent by the fact that in the past, since the UK Vaccination Acts of 1853, imposing fines has been among the tools employed to vaccine refusals (Wolfe and Sharp 2002). Moreover, some governments have already levied financial penalties on SARS-CoV-2 vaccine decliners (see Yoon 2012; Rich and Hida 2021). Hence, it is pivotal to investigate if there are overlooked reasons to question fine-based policy measures.[6] We argue that the tendency to keep options open is one of these reasons.

What counts in human decision processes is not exclusively the expected values assigned to the outcomes. Obviously, the features of the outcomes associated with the available options play a crucial role in human beings' decisions, yet the same can be said about the mere availability of the options themselves, regardless of the attraction exerted by the related outcomes (Simonson 1990; Kahneman and Snell 1992; Walsh 1995; Bown et al. 2003; Carmon et al. 2003). Such attraction does not necessarily lead to desirable choices, indeed could even prevent human beings from optimizing their happiness (Gilbert and Ebert 2002). Jiwoong Shin

[6] Although many concerns have been raised regarding the effectiveness of fine-based policy measures (see Drew 2019), as noted by Gravagna and colleagues, there is a lack of epidemiological studies on the efficacy of mandatory policies: "Rigorous and comprehensive studies that can evaluate which aspects of vaccine mandates, if any, and which types of penalties, if any, are effective at increasing vaccination coverage in multiple contexts are needed" (Gravagna et al. 2020: 7872).

and Dan Ariely conceived a straightforward strategy to test, through the analysis of a series of economic decisions, whether human beings are willing to suffer an economic loss rather than losing their chance to keep their options available (Shin and Ariely 2004). In line with the wording proposed in their work, let's refer to this specific tendency as "keeping doors open", henceforth KDO. Their experimental setting works as follows. Three doors, one blue, one green and the other red, appear on a computer screen. Clicking on a door gives access to a related room. The participants' task consists of distributing a limited number of clicks over the doors (door-clicks) and over the related rooms (room-clicks). After each click, regardless of whether used on a door or on a room, the number of clicks available is reduced by one (Figure 1). A crucial aspect of the experimental set-up is that the room-clicks are associated with actual economic gains, whereas the door-clicks do not guarantee any gain. How much a participant can earn by each room-click strongly depends on which room is entered. Indeed, if on the one hand, the average of the payoff distributions is the same for all the three rooms, on the other hand the skewness and the variance of the payoff distributions differ from door to door, making each option diverse from the alternatives. Due to the fact that door-clicks do not guarantee any gain and that the number of the clicks is fixed, each door-click, with the exception of the first, implies an opportunity cost of 3 ¢ , namely the expected value of each room-click.

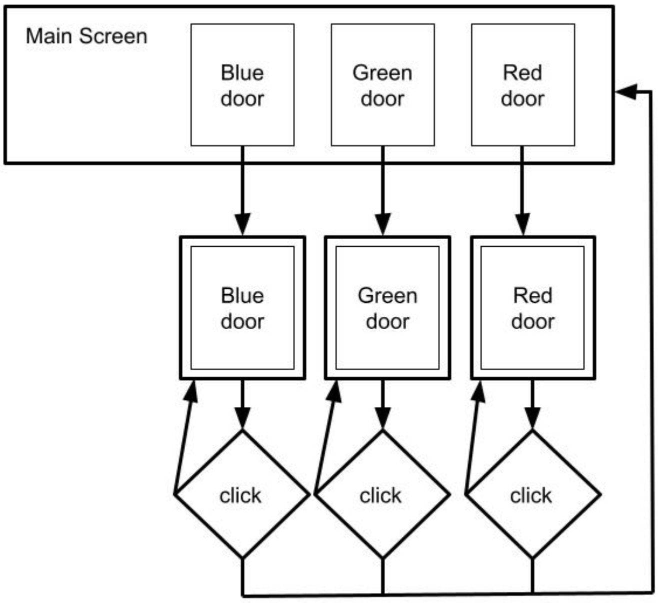


Figure 1

The main possible manipulation in the experiment concerns the consequences of a series of room-clicks. According to the so-called "constant-availability conditions" (henceforth CAC), all the three doors remain available, regardless of how participants decide to distribute the clicks at their disposal through the task. In contrast, the "decreased-availability conditions" (henceforth DAC) involves the eventuality that doors disappear and such eventuality depends directly on the choices made by the participants. Indeed, if the DAC is in place, when a room remains unclicked for 15 consecutive clicks, it disappears permanently. Participants can prevent the disappearance of a door and its related room exclusively by:

1. clicking on the door;
2. spending the subsequent click on the related room.

For instance, the DAC implies that if a participant clicked for the 15th time in a row on the blue room, (s)he would necessarily cause the permanent disappearance of both the red and the green doors. In contrast, in the CAC none of the doors are at risk to disappear. Shin and Ariely have conceived and tested several variations of both these conditions. The results obtained when three variations of the DAC concerning the information availability were in place are particularly relevant for the purpose of the present work. In one variation, participants did not receive any piece of information concerning the payoff distributions associated with each door. This is the "no-prior-information" variation. On the other hand, in the "practice-information" variation, before taking part in the proper and paid game, participants had the chance to practice with the task through a session without financial consequences, being informed that the payoff distributions associated with each room reflect the rooms' payoff distributions of the proper and paid game. Given the third and last variation—the "descriptive-information" variation—participants were explicitly informed about some of the salient features of the payoff distributions right before taking part in the game. In the latter setting, participants were made aware of the fact that the average payoff distribution is the same for each of the three rooms and about the difference in terms of skewness and variance between the rooms' payoff distribution. However, according to this variation, participants were unaware of which skewness and variance of distributions featured which of the three rooms.

Shin and Ariely then measured the mean with which a kind of very peculiar behavior occurs, viz. once a participant has already entered a room (s)he switches to another room, clicks on the corresponding door and, finally, switches back to the previous room. The authors of the article defined such behavior as "pecking". There are two main reasons why participants could decide to switch rooms when the DAC is in place, regardless of the pieces of information available. On the one hand, participants could want to increase their own knowledge of the payoff distributions of each room, perfectible under each variation. On the other hand, switching could be intended to avoid one of the doors permanently disappearing. However, pecking cannot be interpreted as an attempt to gain information on the payoff distributions, in that one room-click provides little information about the payoff distribution associated with the room. Therefore, pecking behavior cannot be prompted by anything but the will to keep options open. The results of the experiment are somewhat surprising: when respondents faced the DAC, the showed a stronger tendency of pecking than when doors were constantly available, irrespective of what variation in terms of information availability was in place. Hence, not only is the KDO tendency in fact triggered by the threat of option-disappearance, but it is also resilient no matter what pieces of information are possessed by the decision-makers concerning the available options and the relative outcomes.

At this point, we suggest that the DAC, given both the practice-information and the descriptive-information variations, depicts a choice environment worryingly analogous to the one inhabited by those who face the vaccine choice in the event that a fine-based compulsory policy is introduced. Let us develop the analogy. Concerning the first aspect of the analogy, the pecking behavior and the payment of a fine due to the refusal of the vaccine are two similar decisions. To see why, let's consider the following scenario that roughly describes what would happen if a SARS-CoV-2 vaccine mandatory policy based on fines was introduced forthwith:

Fine Scenario

The national health system of a state is in possession of several brands of vaccine, about which information circulates freely, like the information on the SARS-CoV-2 virus. Nevertheless, being in short supply, the national health system cannot guarantee a vaccination campaign that rapidly covers all the citizens who can safely be vaccinated. Hence, the government decides which categories of the population to treat first, following certain priorities. Furthermore, for organizational reasons, the government denies citizens the possibility of expressing their preference on which Covid-19 vaccine they will get. In order to maximize the vaccine uptake rate in each category, the government opts to impose monetary weak punishment to those summoned citizens who decline to get vaccinated when their turn occurs. The measure adopted establishes that each time citizens decline to be vaccinated, they incur a fine, the amount of which is always the same irrespective of the number of refusals already made.

This measure would force citizens to choose between two options, on the one hand to get vaccinated and avoid incurring a fine, on the other hand to persist in remaining unvaccinated by paying a fine each time a chance to be inoculated appears.

A salient feature of the *Fine Scenario* is alarmingly similar to a feature of the DAC described above. Indeed, in both cases the decision-makers are asked to perform a costly act to avoid the disappearance of one of the options available from the beginning. In the DAC the respondent who wants to keep the doors open has to undertake the opportunity cost of the door-click spent on the disappearing door. Likewise, in the vaccine scenario the citizens are asked to pay a fine to avert the possibility that one of the options, namely to remain unvaccinated, definitively disappears. So, in both cases, keeping the options open implies incurring an economic loss.

Moreover, there is a second aspect for which the choice environments that characterize the DAC and the *Fine Scenario* are analogous. This aspect concerns the partiality of the information available and the uncertainty about the possible outcomes of the decision. In the DAC, knowledge of the gains obtainable by clicking the three rooms is perfectible, regardless of which variation of information availability is in place. In the "no-prior-information" this is obviously the case, but the same goes for the "practice-information" variation, since the practice session does not last enough to get enough information on the payoff distributions. Furthermore, in the "descriptive-information" variation, participants are unaware of which distribution skewness and variance featured which of the three rooms and so even their level of information can be improved. Likewise, in the *Fine Scenario*, the information on both the SARS-CoV-2 virus and vaccine brands, which circulates freely among citizens, leads to the citizens' partial uncertainty about the outcomes. Citizens have easy and constant access to a dense flow of information, all relevant for the vaccine choice. Scientific journals, mass media, social media, and circles of acquaintances are producing a massive quantity of information on COVID-19 and SARS-CoV-2 vaccines, to the point that it has led scholars to define the situation as an infodemic (Gallotti et al. 2020). Citizens are having constant access to an impressive amount of information on the features that characterize the options at hand, especially on the possible side effects related to the uptaking of SARS-CoV-2 vaccines, the chance to contract the virus, and the related health consequences. The point is that such pieces of information define a condition of partial uncertainty exactly like the condition experienced by the participants assigned to all the three versions of DAC.

In the light of these analogies between the experimental setting conceived by Shin and Ariely and the *Fine Scenario* just outlined, it is reasonable to expect that the KDO tendency would play a not negligible role on the vaccine choice, whether or not a fine-based coercive policy was introduced. Furthermore, the KDO tendency could emerge regardless of the degree of information obtained, the nature of the resources exploited and the tenability of the pieces of information collected by citizens. As said in §2, governments of liberal democracies can neither adopt policies in which citizens are compelled by force to get the vaccine nor impose fines on vaccine decliners high enough to overwhelm the expected gain of refusing the vaccine for the vast majority of the citizens. Hence, a fine-based mandatory policy scenario, whereby the punishments imposed are weak, is a policy that could be realistically implemented by a government of a liberal democracy to address the COVID-19 emergency. Unfortunately, the human KDO tendency leads to the conjecture that the fines imposed within a policy so characterized could be perceived as an opportunity to keep options open rather than as a deterrent by those citizens who have some interest in both avoiding vaccination and being vaccinated. Strong vaccine refusers do not have any interest in leaving open the opportunity to get vaccinated, hence they perceive the weak financial disincentive described in the *Fine Scenario* as a cost to take on rather than a tempting opportunity. So, to claim that the KDO tendency plays a role in strong vaccine refusers' choices is hardly plausible.

Rather, we specifically refer to the phenomenon called "vaccine hesitancy", which is one of the main concerns among the governments engaged in SARS-CoV-2 vaccines campaigns (Lazarus et al. 2020). Vaccine hesitant are those who "[...] delay in acceptance or refusal of vaccines despite availability of vaccine services" (SAGE 2014: 7) and so arguably the category attracted to a mean that will allow them to persist in their hesitancy.[7] Concerning the COVID-19 emergency, vaccine hesitancy especially raises concerns because one of its determinants is the exposition of conspiracy theories (Jolley and Douglas 2014; Roozenbeek et al. 2020) which are rapidly spreading through social media nowadays. To make things even worse, it seems reasonable to believe that concerning the COVID-19 emergency, introducing such a policy would trigger the KDO tendency in vaccine hesitant to an even larger extent than in the DAC described by Shin and Ariely (2004). The two cases at hand are indeed typically dissimilar in terms of preference ratio. When the DAC was in place, the average of pecking behaviors was high, despite the fact that the disappearing doors were related

[7] For a more detailed analysis of the vaccine hesitancy phenomenon that captures the difference between denialism and refusal, see Navin 2015.

to rooms of little interest for the chooser, as shown by previous decisions. To the contrary, concerning the SARS-CoV-2 vaccines, vaccine hesitant assigns a considerable interest to avoiding vaccination or to get an alternative vaccine than the assigned one, interests sufficient to make them hesitant about what to do. This is particularly worrying from a policy making standpoint in that vaccine hesitant should be the main target of policy measures, being the most sensitive to changes in the choice environment and representing a vast part of the population (Leask 2011; Buttenheim 2020). Hence, policies that impose weak financial disincentives on vaccine decliners seem to provide an especially fertile ground to lead to interpret fines as unmissable opportunities to keep options open.

4. A Laboratory Experiment

In the present manuscript we initially considered peer pressure, the default effect, and case versus base-rate effect as examples of psychological mechanisms relevant for the vaccine choice as well some nudges that can be used to moderate these. Then we focused on what B&C scientists have investigated regarding the conditions under which imposing fines turns out to be an effective policy measure. In particular, we discussed the case of the crowding-out effect as a reason to be skeptical about the introduction of fines as a way to curb vaccine declination. The major contribution of the paper is our claim that there is another and somehow neglected reason to be skeptical, i.e. the KDO tendency, which is threatening in light of the vaccine hesitancy phenomena.

Nevertheless, that imposing a fine could be perceived by vaccine hesitant as an unmissable opportunity to keep the options open rather than a deterrent to hesitancy is so far nothing more than a reasonable conjecture. Hence, at this stage, our claim has only limited practical use for policy-makers. We suggest a design for testing our conjecture. In exploring its reliability for policy purposes, a laboratory experiment could be set as follows. Participants could be randomly assigned to one of two conditions. Those assigned to the first condition would be asked to read a scenario in which the government of his/her home state introduced a coercive policy measure, imposing fines on the citizens who decline to be vaccinated when summoned. In this condition, the citizens would have several chances to be vaccinated, let's say one month away from each other. This condition mirrors the *Fine Scenario*. The second condition, then, could depict a more tragic emergency whereby the home state of the participant is said to be suffering from a really severe short in vaccine supply due to some kind of force majeure, for instance vaccine batches repetitively getting lost due to natural disasters or terrorist attacks. These setbacks will have made it impossible for the health system to guarantee further chances to get vaccinated in the near future for those who refuse the vaccine when summoned for the first time. Indeed, the second chance will not come before a considerable amount of time, let's say at least two years succeeding the first chance. Like in the first scenario, refusing to be vaccinated is penalized with a fine. In this last condition, to refuse to get the vaccine jab when the first chance comes means to miss out on the vaccination coverage for an extended part of the COVID-19 emergency, which equates to the disappearance of the option of being vaccinated, although not permanently, arguably at the time of greatest need. Conversely, in the first condition the vaccine refuser finds in the fine an opportunity to keep both the options open during the COVID-19 emergency. Asking the participants assigned to each condition about their propensity to refuse the vaccine when their first turn comes is a preliminary way to test if the KDO tendency is a determinant. Indeed, if this is the case, the propensity to refuse the vaccine should be higher given the first condition than the second one, in that only the first condition entails a policy in which the fine imposed allows one to keep both the options open over the course of the pandemic. The results from the experiment just outlined would be a good first step to investigate if the KDO tendency has in fact a role in influencing the vaccine choice. However, considering the eventual results of the experiment as a research result ready to be taken into account in planning policies on the vaccine uptake should be firmly avoided. The reason for such cautious approach is the subject of our concluding remarks.

5. Concluding Remarks

We share the position defended by many scholars that B&C scientists should be extremely careful in considering the research results from their fields as pieces of knowledge relevant for facing the COVID-19 emergency (e.g. IJzerman et al. 2020). Indeed, due to methodological drawbacks, the solidity and the quality of the vast majority of the B&C research potentially relevant for anti-COVID-19 policies are not good enough to play a role in policy decision processes. In fact, overall the psychological sciences are going through a reproducibility crisis (Open Science Collaboration 2015), and research from B&C sciences often relies on samples drawn from a slice of populations that is scarcely representative (Henrich et al. 2010). Furthermore, these research investigations are often based in laboratory settings and aim to collect data on intention through self-reported scales which, in both cases, provide results that say little about actual behavior (Webb & Sheeran 2006; Baumeister et al. 2007; Levitt and List 2007). Finally, often the very same phenomenon can clearly emerge in one context but not, or just in a milder form, in a different one (Shimizu and Udagawa 2018). It could be the case of the KDO tendency, which has been repetitively found to influence economic choices, could be mild or even absent when human lives are at stake.

All of this should lead B&C scientists toward *epistemic humility* regarding the role that should be played by the investigative insights from their fields in shaping anti-COVID policies. The epistemic humility can be waived when the results obtained concern reversible kinds of events as, for instance, in the case of the strength of peer pressure to dissuade someone from littering (Cialdini et al., 1990). Nevertheless, when the events are irreversible, epistemic humility is mandatory. Clearly, this is the case with the COVID-19 emergency, where the health and lives of the world population are at stake. In other words, the risks of over-generalizing could not be higher. Hopefully, this risk is well understood by B&C scientists, 681 of which signed in 2020 a public letter to distance themselves from the UK government. These B&C scientists asked the UK government to avoid the mistake of considering "behavioral fatigue" as a justification for delaying the introduction of strict measures to fight against COVID-19, being "not convinced that enough is known about 'behavioural fatigue' or to what extent these insights apply to the current exceptional circumstances" (Open Letter 2020).[8]

Other than advocating caution, B&C scientists are taking action to improve the robustness of their research results to make them ready for policy-makers. Firstly, projects like the Psychological Science Accelerator (PSA) have emerged to improve the reliability of experimental results and test the generalizability of insights from psychological science through crowdsourced research carried on by networks of laboratories (Moshontz et al. 2018). Secondly, the need for science-based insights to address the COVID-19 emergency has spurred B&C scientists to focus on the methodological issues of their disciplines. Currently, B&C scientists are discussing the criteria that research should fulfill for results to be considered suitable to use to plan-COVID policies. Right after the COVID-19 pandemic was declared, two main criteria were advanced to assess the readiness of the research results from B&C sciences. Kai Ruggeri and colleagues proposed a framework called "Theoretical, Empirical, Applicable, and Replicable Impact rating system" (THEARI) based on a five-tier rating. This system ranges from a pure theoretical level to the highest level in which the findings are "validated at the highest conceivable power (i.e., populations) through real-world testing and replication of effects in multiple settings" (Ruggeri et al. 2020). Furthermore, Hans IJzerman and colleagues proposed a rating system of nine levels of evidence readiness adapted from the system NASA uses to assess the maturity level of technologies (IJzerman et al. 2020). The experiment we sketched at the beginning of this section would be at the lower levels of both these scales, hence categorizing it as unready to be used in making policy decisions. However, as it is reasonable to believe that the KDO tendency plays a role in vaccine choice, we believe it is worth it to begin testing it. If it turns out it does have validity, B&C scientists could advance the research to the point where it could be suitable for policy implementation. Possibly, the results obtained could make it urgent to investigate the connection between each of the determinants of a complex phenomenon as vaccine hesitancy and the KDO tendency. The determinants of vaccine hesitancy are indeed different in nature from each other, being related to negative emotions triggered by vaccines, bounded rationality, misinformation, and as well the introduction of information, as the availability of a new vaccine (SAGE 2014). For this reason, it might turn out that some determinants and groups of them affect the KDO tendency differently than others.[9]

[8] At the beginning of the COVID-19 emergency, the UK government preferred a mild "keep calm and carry on" approach. Only on the 23rd of March 2020, slow off the mark, did Prime Minister Boris Johnson make an unavoidable U-turn and impose more strict measures. British mass media have imputed part of the responsibility for the delay in the intervention to some of the government's behavioral consultants. Allegedly, some members of the "Behavioural Insights Team" brought "behavioural fatigue" into play as a reason to delay the introduction of strict measures (Yates 2020).

[9] The authors would like to express their gratitude to two anonymous reviewers for their constructive and in-depth comments on this article.

References

Abel, T. and McQueen, D. 2020, "The COVID-19 Pandemic Calls for Spatial Distancing and Social Closeness: Not for Social Distancing!", *International Journal of Public Health*, 65, 231, doi: 10.1007/s00038-020-01366-7.

Abrams, D., Wetherell, M., Cochrane, S., Hogg, M.A., and Turner, J.C. 1990, "Knowing What to Think by Knowing Who You Are: Self-Categorization and the Nature of Norm Formation, Conformity and Group Polarization", *British Journal of Social Psychology*, 29, 97-119, doi: 10.1111/j.2044-8309.1990.tb00892.x.

Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. 2009, "Large Stakes and Big Mistakes", *Review of Economic Studies*, 76, 451-69, doi: 10.1111/j.1467-937x.2009.00534.x.

Baumeister, R.F., Vohs, K.D., and Funder, D.C. 2007, "Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior?", *Perspectives on Psychological Science*, 2, 396-403, doi : 10.1111/j.1745-6916.2007.00051.x.

Bavel, J.J.V., Baicker, K., Boggio, S. *et al.* 2020, "Using Social and Behavioural Science to Support COVID-19 Pandemic Response", *Nature Human Behaviour*, 4, 460-71, doi: 10.1038/s41562-020-0884-z.

Becker, G. 1968, "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, 76, 169-217, www.jstor.org/stable/1830482.

Betsch, C., Böhm, R., and Korn, L. 2013, "Inviting Free-riders or Appealing to Prosocial Behavior? Game-theoretical Reflections on Communicating Herd Immunity in Vaccine Advocacy", *Health Psychology*, 32, 978-85, doi: 10.1037/a0031590.

Bicchieri, C. 2016, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*, Oxford: Oxford University Press.

Bicchieri, C., and Dimant, E. 2019, "Nudging with Care: The Risks and Benefits of Social Information", *Public Choice*, doi: 10.1007/s11127-019-00684-6.

Bicchieri, C., Dimant, E., and Xiao, E. 2021, "Deviant or Wrong? The Effect of Norm Information on the Efficacy of Punishment", *SSRN Electronic Journal*, doi: 10.2139/ssrn.3779018.

Bicchieri, C., Fatas, E., Aldama, A., Casas, Deshpande, I., Lauro, M., Parilli, C., Spohn, M., Pereira, P., and Wen, R. 2020, "In Science we (Should) Trust: Expectations and Compliance During the COVID-19 Pandemic", doi: 10.21203/rs.3.rs-106840/v1.

Bish, A., Yardley, L., Nicoll, A., and Michie, S. 2011, "Factors Associated with Uptake of Vaccination Against Pandemic Influenza: A Systematic Review", *Vaccine*, 29, 6472-84, doi: 10.1016/j.vaccine.2011.06.107.

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., and Fowler, J.H. 2012, "A 61-million-person Experiment in Social Influence and Political Mobilization", *Nature*, 489, 295-98, doi: 10.1038/nature11421.

Bowles, S. 2008, "Policies Designed for Self-Interested Citizens May Undermine "The Moral Sentiments": Evidence from Economic Experiments", *Science*, 320, 1605-1609, doi: 10.1126/science.1152110.

Bown, N.J., Read, D., and Summers, B. 2003, "The Lure of Choice", *Journal of Behavioral Decision Making*, 16, 297-308, doi: 10.1002/bdm.447.

Bruine de Bruin, W., Parker, A.M., Galesic, M., and Vardavas, R. 2019, "Reports of Social Circles' and Own Vaccination Behavior: A National Longitudinal Survey", *Health Psychology*, 38, 975-83, doi: 10.1037/hea0000771.

Buttenheim, A.M. 2020, "SARS-CoV-2 Vaccine Acceptance: We May Need to Choose Our Battles", *Annals of Internal Medicine*, 173, 1018-19, doi: 10.7326/m20-6206.

Buttenheim, A.M., and Asch, D.A. 2013, "Making Vaccine Refusal Less of a Free Ride", *Human Vaccines & Immunotherapeutics*, 9, 2674-75, doi: 10.4161/hv.26676.

Calabuig, V., Fatas, E., Olcina, G., and Rodriguez-Lara, I. 2016, "Carry a Big Stick, or No Stick at all", *Journal of Economic Psychology*, 57, 153-71, doi: 10.1016/j.joep.2016.09.006.

Carmon, Z., Wertenbroch, K., and Zeelenberg, M. 2003, "Option Attachment: When Deliberating Makes Choosing Feel like Losing", *Journal of Consumer Research*, 30, 15-29, doi: 10.1086/374701.

Chapman, G.B., Li, M., Colby, H., and Yoon, H. 2010, "Opting In vs Opting Out of Influenza Vaccination", *JAMA*, 304, 43-44, doi: 10.1001/jama.2010.892.

Cialdini, R.B., and Goldstein, N.J. 2004, "Social Influence: Compliance and Conformity", *Annual Review of Psychology*, 55, 591-621, doi: 10.1146/annurev.psych.55.090902.142015.

Cialdini, R.B., Reno, R.R., and 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, 1015-26, doi: 10.1037/0022-3514.58.6.1015.

Drew, L. 2019, "The Case for Mandatory Vaccination", *Nature*, 575, 58-60, doi: 10.1038/d41586-019-03642-w.

European Parliament and European Council, 15 June 2021, Regulation 2021/953 on a framework for the issuance, verification and acceptance of interoperable COVID-19 vaccination, test and recovery certificates (EU Digital COVID Certificate) to facilitate free movement during the COVID-19 pandemic, L 211/1.

Fehr, E. and Falk, A. 2002, "Psychological Foundations of Incentives", *European Economic Review*, 46, 687-724, doi: 10.1016/s0014-2921(01)00208-2.

Fehr, E. and Gächter, S. 2002, "Do Incentive Contracts Undermine Voluntary Cooperation?", *SSRN Electronic Journal*, doi: 10.2139/ssrn.313028.

Fehr, E. and Rockenbach, B. 2003, "Detrimental Effects of Sanctions on Human Altruism", *Nature*, 422, 137-40, doi: 10.1038/nature01474.

Felletti, S. 2020, "'Trust me, I'm your Neighbour'. How to Improve Epidemic Risk Containment through Community Trust", *Mind & Society*, 1, 4, doi: 10.1007/s11299-020-00266-w.

Fischbacher, U., Gächter, S., and Fehr, E. 2001, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", *Economics Letters*, 71, 397-404, doi: 10.1016/s0165-1765(01)00394-9.

Gallotti, R., Valle, F., Castaldo, N., Sacco, P., and De Domenico, M. 2020, "Assessing the Risks of 'Infodemics' in Response to COVID-19 Epidemics", *Nature Human Behaviour*, 4, 1285-93, doi: 10.1038/s41562-020-00994-6.

Gilbert, D.T. and Ebert, J.E.J. 2002, "Decisions and Revisions: The Affective Forecasting of Changeable Outcomes", *Journal of Personality and Social Psychology*, 82, 503-14.

Gneezy, U., Meier, S., and Rey-Biel, P. 2011, "When and Why Incentives (Don't) Work to Modify Behavior", *Journal of Economic Perspectives*, 25, 191-210, doi: 10.1257/jep.25.4.191.

Gneezy, U. and Rustichini, A. 2000, "A Fine is a Price", *The Journal of Legal Studies*, 29, 1-17.

Goldstein, N.J., Cialdini, R.B., and Griskevicius, V. 2008, "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels", *Journal of Consumer Research*, 35, 472-82, doi: 10.1086/586910.

Goodwin, T. 2012, "Why We Should Reject 'Nudge'", *Politics*, 32, 85-92.

Gravagna, K., Becker, A., Valeris-Chacin, R., Mohammed, I., Tambe, S., Awan, F. A., Toomey, T.L., and Basta, N.E. 2020, "Global Assessment of National Mandatory Vaccination Policies and Consequences of Non-compliance", *Vaccine*, 38, 7865-73, doi: 10.1016/j.vaccine.2020.09.063.

Grüne-Yanoff, T. 2012, "Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles", *Social Choice and Welfare*, 38, 635-45.

Halpern, S.D., Ubel, P.A., and Asch, D.A. 2007, "Harnessing the Power of Default Options to Improve Health Care", *New England Journal of Medicine*, 357, 1340-44, doi: 10.1056/nejmsb071595.

Hayes, K.J., Eljiz, K., Dadich, A., Fitzgerald, J.A., and Sloan, T. 2015, "Trialability, Observability and Risk Reduction Accelerating Individual Innovation Adoption Decisions", *Journal of Health Organization and Management*, 29, 271-94, doi: 10.1108/jhom-08-2013-0171.

Henrich, J., Heine, S.J., Norenzayan, A. 2010, "The Weirdest People in the World?", *Behavioral and Brain Sciences*, 33, 61-83.

IJzerman, H., Lewis, N.A., Przybylski, A.K., Weinstein, N., DeBruine, L., Ritchie, S.J., Vazire, S., Forscher, P.S., Morey, R.D., Ivory, J.D., and Anvari, F. 2020, "Use Caution When Applying Behavioural Science to Policy", *Nature Human Behaviour*, 4, 1092-94, doi: 10.1038/s41562-020-00990-w.

Jachimowicz, J., Duncan, S., Weber, E., and Johnson, E. 2019, "When and Why Defaults Influence Decisions: A Meta-analysis of Default Effects", *Behavioural Public Policy*,3, 159-86, doi: 10.1017/bpp.2018.43.

Johnson, E.J., and Goldstein, D. 2003, "Do Defaults Save Lives?", *Science*, 302, 1338-39, doi: 10.1126/science.1091721.

Jolley, D., and Douglas, K. M. 2014, "The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions", *PLoS ONE*, 9, e89177, doi: 10.1371/journal.pone.0089177.

Kahneman, D., Knetsch, J.L., and Thaler, R.H. 1990, "Experimental Tests of the Endowment Effect and the Coase Theorem", *Journal of Political Economy*, 98, 1325-48.

Kahneman, D. and Snell, J. 1992, "Predicting a Changing Taste: Do People Know What They Will Like?", *Journal of Behavioral Decision Making*, 5, 187-200, doi: 10.1002/bdm.3960050304.

Korn, L., Betsch, C., Böhm, R., and Meier, N.W. 2018, "Social Nudging: The Effect of Social Feedback Interventions on Vaccine Uptake", *Health Psychology*, 37, 1045-54, doi: 10.1037/hea0000668.

Lazarus, J.V., Ratzan, S.C., Palayew, A., Gostin, L.O., Larson, H.J., Rabin, K. et al. 2020, "A Global Survey of Potential Acceptance of a COVID-19 Vaccine", *Nature Medicine*, doi: 10.1038/s4159 1-020-1124-9.

Leask, J. 2011, "Target the Fence-Sitters", *Nature*, 473, 443-45, doi: 10.1038/473443a.

Lehmann, B.A., Chapman, G.B., Franssen, F.M., Kok, G., and Ruiter, R.A. 2016, "Changing the Default to Promote Influenza Vaccination Among Health Care Workers", *Vaccine*, 34, 1389-92, doi: 10.1016/j.vaccine.2016.01.046.

Levine-Tiefenbrun, M., Yelin, I., Katz, R. et al. 2021, "Decreased SARS-CoV-2 Viral Load Following Vaccination", doi: 10.1101/2021.02.06.21251283.

Levitt, S. and List, J. 2007, "What Do Laboratory Experiments Tell Us About the Real World?", *Journal of Economic Perspectives*, 21, 153-74.

Lunn, P.D., Belton, C.A., Lavin, C., McGowan, F.P., Timmons, S., and Robertson, D.A. 2020, "Using Behavioral Science to Help Fight the Coronavirus", *Journal of Behavioral Public Administration*, 3, doi: 10.30636/jbpa.31.147.

Lynch, J.G., and Ofir, C. 1989, "Effects of Cue Consistency and Value on Base-rate Utilization", *Journal of Personality and Social Psychology*, 56, 170-81, doi: 10.1037/0022-3514.56.2.170.

Madrian, B.C. 2014, "Applying Insights from Behavioral Economics to Policy Design", *Annual Review of Economics*, 6, 663-88, doi: 10.1146/annurev-economics-080213-041033.

Magnano, R. 2019, "Vaccini, Fuori Chi non è in Regola ma le Sanzioni Restano sulla Carta", *Il Sole 24 Ore*, www.ilsole24ore.com/art/vaccini-fuori-chi-non-e-regola-ma-sanzioni-restano-carta-ABbYnucB.

Mallapaty, S. 2021, "Can COVID Vaccines Stop Transmission? Scientists Race to Find Answers", *Nature*, doi: 10.1038/d41586-021-00450-z.

Maringe, C., Spicer, J., Morris, M., Purushotham, A., Nolte, E., Sullivan, R., Rachet, B., and Aggarwal, A. 2020, "The Impact of the COVID-19 Pandemic on Cancer Deaths Due to Delays in Diagnosis in England, UK: A National, Population-Based, Modelling Study", *The Lancet Oncology*, 21, 1023-34, doi: 10.1016/s 1470-2045(20)30388-0.

McCormick, K.D., Jacobs, J.L., and Mellors, J.W. 2021, "The Emerging Plasticity of SARS-CoV-2", *Science*, 371, 1306-08, doi: 10.1126/science.abg4493.

Mellström, C. and Johannesson, M. 2008, "Crowding Out in Blood Donation: Was Titmuss Right?", *Journal of the European Economic Association*, 6, 845-63, doi: 10.1162/jeea.2008.6.4.845.

Ministry of Health of Israel (n.d.), What is a Green Pass?, Retrieved March 28 2021, from corona.health.gov.il/en/directives/green-pass-info/.

Moshontz, H., Campbell, L., Ebersole, C.R., IJzerman, H., Urry, H.L., Forscher, P.S. et al. 2018, "The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network", *Advances in Methods and Practices in Psychological Sciences*, 1, 501-15, doi: 10.1177/2515245918797607.

Nakayachi, K., Ozaki, T., Shibata, Y., and Yokoi, R. 2020, "Why Do Japanese People Use Masks Against COVID-19, Even Though Masks Are Unlikely to Offer Protection from Infection?", *Frontiers in Psychology*, 11, doi: 10.3389/fpsyg.2020.01918.

Navin, M. 2016, *Values and Vaccine Refusal: Hard Questions in Ethics, Epistemology, and Health Care*, New York: Routledge.

Open Letter (2020, March 16), *Open letter to the UK Government regarding COVID-19*. Retrieved March 28 2021, from sites.google.com/view/covidopenletter/home.

Open Science Collaboration 2015, "Estimating the Reproducibility of Psychological Science", *Science*, 349, 13, doi: 10.1126/science.aac4716.

Petter, E., Mor, O., Zuckerman, N., Oz-Levi, D., Younger, A., Aran, D., and Erlich, Y. 2021, "Initial Real World Evidence for Lower Viral Load of Individuals Who Have Been Vaccinated by BNT162b2", doi: 10.1101/2021.02.08.21251329.

Phelan, A.L. 2020, "COVID-19 Immunity Passports and Vaccination Certificates: Scientific, Equitable, and Legal Challenges", *The Lancet*, 395, 1595-98, doi: 10.1016/s0140-6736(20)31034-5.

Quadri-Sheriff, M., Hendrix, K.S., Downs, S.M., Sturm, L.A., Zimet, G.D., and Finnell, S.M.E. 2012, "The Role of Herd Immunity in Parents' Decision to Vaccinate Children: A Systematic Review", *Pediatrics*, 130, 522-30, doi: 10.1542/peds.2012-0140.

Rich, M. and Hida, H. 2021, "As the Pandemic Took Hold, Suicide Rose among Japanese Women", *The New York Times*, www.nytimes.com/2021/02/23/world/as-the-pandemic-took-hold-suicide-rose-among-japanese-women.html.

Roozenbeek, J., Schneider, C.R., Dryhurst, S., Kerr, J., Freeman, A.L.J., Recchia, G., van der Bles, A.M., and van der Linden, S. 2020, "Susceptibility to Misinformation about COVID-19 around the World", *Royal Society Open Science*, 7, 201199, doi: 10.1098/rsos.201199.

Ruggeri, K., van der Linden, S., Wang, Y.C., Papa, F., Riesch, J., and Green, J. 2020, "Standards for Evidence in Policy Decision-making", *Nature Research Social and Behavioural Sciences*, 399005, go.nature.com/2zdTQIs.

Strategic Advisory Group of Experts on Immunization (SAGE) 2014, *Report of the SAGE Working Group on Vaccine Hesitancy*, Retrieved September 12 2021, from www.who.int/immunization/sage/meetings/2014/october/SAGE_working_group_revised_report_vaccine_hesitancy.pdf?ua=1.

Santos Rutschman, A. 2020, "Why the Government Shouldn't Pay People to Get Vaccinated Against COVID-19", *SSRN Electronic Journal*, doi: 10.2139/ssrn.3740198.

Schmelz, K. 2020, "Enforcement May Crowd Out Voluntary Support for COVID-19 Policies, Especially Where Trust in Government is Weak and in a Liberal Society", *Proceedings of the National Academy of Sciences*, 118, e2016385118, doi: 10.1073/pnas.2016385118.

Schmidt, A. 2017, "The Power to Nudge", *American Political Science Review*, 111, 404-17.

Schultz, P.W., Nolan, J.M., Cialdini, R.B., Goldstein, N.J., and Griskevicius, V. 2007, "The Constructive, Destructive, and Reconstructive Power of Social Norms", *Psychological Science*, 18, 429-34, doi: 10.1111/j.1467-9280.2007.01917.x.

Shimizu, K. and Udagawa, D. 2018, "Is Human Life Worth Peanuts? Risk Attitude Changes in Accordance with Varying Stakes", *Plos One*, 13, e0201547, doi: 10.1371/journal.pone.0201547.

Shin, J. and Ariely, D. 2004, "Keeping Doors Open: The Effect of Unavailability on Incentives to Keep Options Viable", *Management Science*, 50, 575-86, doi: 10.1287/mnsc.1030.0148.

Siani, A. 2019, "Measles Outbreaks in Italy: A Paradigm of the Re-emergence of Vaccine-Preventable Diseases in Developed Countries", *Preventive Medicine*, 121, 99-104, doi: 10.1016/j.ypmed.2019.02.011.

Signorelli, C., Iannazzo, S., and Odone, A. 2018. "The Imperative of Vaccination Put into Practice", *Lancet Infectious Diseases*, 18, 26-27, doi: 10.1016/S1473-3099 (17)30696-5.

Simonson, I. 1990, "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior", *Journal of Marketing Research*, 27, 150, doi: 10.2307/3172842.

Sparkman, G. and Walton, G.M. 2017, "Dynamic Norms Promote Sustainable Behavior, Even If It Is Counternormative", *Psychological Science*, 28, 1663-74, doi: 10.1177/0956797617719950.

Subbaraman, N. 2021, "How Do Vaccinated People Spread Delta? What the Science Says", *Nature*, 596, 327-28, doi: 10.1038/d41586-021-02187-1.

Teitelbaum, J.C. and Zeiler, K. 2018, *Research Handbook on Behavioral Law and Economics (Research Handbooks in Law and Economics series)*, Cheltenham: Edward Elgar Publishing.

Thaker, J. 2020, "Planning for a COVID-19 Vaccination Campaign: The Role of Social Norms, Trust, Knowledge, and Vaccine Attitudes", Preprint 10.31234/ osf.io/q8mz6.

Thaler, R.H. 2021, "Getting Everyone Vaccinated, With 'Nudges' and Charity Auctions", *The New York Times*, www.nytimes.com/2020/12/09/business/corona virus-vaccination-auctions-celebrities.html.

Thaler, R.H. and Sunstein, C. 2008, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, New Haven: Yale University Press.

Trentini, F., Poletti, P., Melegaro, A., and Merler, S. 2019, "The Introduction of 'No jab, No school' Policy and the Refinement of Measles Immunisation Strategies in High-income Countries", *BMC Medicine*, 17, doi: 10.1186/s12916-019-1318-5.

Tversky, A., and Kahneman, D. 1973, "Availability: A Heuristic for Judging Frequency and Probability", *Cognitive Psychology*, 5, 207-32, doi: 10.1016/0010-0285 (73)90033-9.

Walkinshaw, E. 2011, "Mandatory Vaccinations: The International Landscape", *Canadian Medical Association Journal*, 183, E1167-E1168, doi: 10.1503/cmaj.109-3993.

Walsh, J.W. 1995, "Flexibility in Consumer Purchasing for Uncertain Future Tastes", *Marketing Science*, 14, 148-65, doi: 10.1287/mksc.14.2.148.

Webb, T.L. and Sheeran, P. 2006, "Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence", *Psychological Bulletin,* 132, 249-68, doi: 10.1037/0033-2909.132.2.249.

Wolfe, R.M. and Sharp, L.K. 2002, "Anti-Vaccinationists Past and Present", *BMJ*, 325, 430-32, doi: 10.1136/bmj.325.7361.430.

Wood, S. and Schulman, K. 2021, "Beyond Politics—Promoting Covid-19 Vaccination in the United States", *New England Journal of Medicine*, 384, e23, doi: 10.1056/nejmms2033790.

World Health Organization 2020, "Behavioural Considerations for Acceptance and Uptake of Covid-19 Vaccines: Who Technical Advisory Group on Behavioural Insights and Sciences for Health", I-Ii, Rep., doi:10.2307/resrep27868.1.

Xiao, E. 2018, "Punishment, Social Norms, and Cooperation", in Teitelbaum, C.J. and Zeiler, K. (eds.), *Research Handbook on Behavioral Law and Economics*, Cheltenham: Edward Elgar Publishing, 155-73.

Xiao, X. and Borah, P. 2020, "Do Norms Matter? Examining Norm-Based Messages in HPV Vaccination Promotion", *Health Communication*, doi: 10.1080/10410236.2020.1770506.

Yates, T. 2020, "Why is the Government Relying on Nudge Theory to Fight Coronavirus?", *The Guardian*, www.theguardian.com/commentisfree/2020/mar/13/why-is-the-government-relying-on-nudge-theory-to-tackle-coronavirus.

Yoeli, E., Hoffman, M., Rand, D.G., and Nowak, M.A. 2013, "Powering Up with Indirect Reciprocity in a Large-Scale Field Experiment", *Proceedings of the National Academy of Sciences*, 110, 10424-429, doi: 10.1073/pnas.1301210110.

Yoon, D. 2021, "Declining a Covid-19 Vaccine Risks Penalties in Some Countries", *The Wall Street Journal*, www.wsj.com/articles/declining-a-covid-19-vaccine-risks-penalties-in-some-countries-11613998997.

Advisory Board

Can a City Be Relocated? Exploring the Metaphysics of Context-Dependency

Fabio Bacchini and Nicola Piras***

** University of Sassari*

*** University of Milan*

Abstract

This paper explores the Persistence Question about cities, that is, what is necessary and sufficient for two cities existing at different times to be numerically identical. We first show that we can possibly put an end to the existence of a city in a number of ways other than by physically destroying it, which reveals the metaphysics of cities to be partly different from that of ordinary objects. Then we focus in particular on the commonly perceived vulnerability of cities to imaginary relocation; and we make the hypothesis that cities do have among their essential properties that of being surrounded by a specific geographical context. Finally, we investigate the possibility that a city can survive relocation in virtue of the capacity of its geographical context to survive it in the first place. We suggest that city contexts may not be essentially context-dependent in turn, and outline a possible description of the criteria for their persistence over time.

Keywords: Persistence, Metaphysics, Essentialism, Context, City.

1. The Persistence Question about Cities

Generally speaking, the Persistence Question is a question about what is necessary and sufficient for two cities existing at different times to be numerically identical. Rising the Persistence Question about cities amounts to asking what is necessary and sufficient for a past (or future) city to exist now.

Indeed, to raise the Persistence Question about cities may appear pointless to many people. Firstly—the objection goes—it is very infrequent that a city may stop to exist. Cities normally persist for a much longer time than people. Moreover, they tend to stop to exist in connection to the collapse of a society, an empire or a nation. But these kinds of events are more and more rare in our global world. Thus, cities (differently from villages) can be supposed to stop to exist in the next future at even a smaller rate than in the past of human history. Secondly, when a city happens to stop to exist, there is no doubt about what has

Argumenta (2021): 217-231
ISSN 2465-2334

DOI 10.14275/2465-2334/20200.bac

happened. In other words, events counting as a city stopping to exist tend to be highly recognisable, in virtue of their necessarily consisting in the physical destruction of the city itself. Thus raising the Persistence Question about cities cannot be a useful philosophical activity.

We rebut that reasoning about the Persistence Question about *x* is always an excellent way to reveal what our concept of *x* is like. In fact, by exploring how *x* survives or not different kinds of change (no matter that some of them are not technically producible), we cast light on the most hidden characteristics of our very conception of *x*, metaphysically speaking. In particular, we can use thought experiments in order to investigate how, according to our best[1] intuitions and judgements, a city can survive some kinds of events and cannot survive others. This discloses what properties are constitutive of a city, and what properties are merely contingent. So thought experiments about cities reveal cities' metaphysical secrets. And increased awareness of the metaphysical nature of a city—how different it is with respect to that of ordinary objects; what kinds of items a city is dependent of; what kinds of items, on the contrary, do not ground its existence—may in turn affect our way of reflecting about cities, as well as governing, planning, bettering, living them.

Of course resorting to using thought experiments to explore how cities can stop to exist may reveal disappointing if a city could stop to exist only as an effect of a physical destruction of all or at least the majority of its parts (buildings, streets, and so on). Yet it seems to us that this is not the case.

2. How We Can Possibly Put an End to the Existence of a City other than by Physically Destroying It

Apparently, we can possibly put an end to the existence of a city other than by physically destroying it. This imaginary exploration may reveal that a city is subject to special persistence conditions that are partially different from those holding for ordinary objects.

A first scenario is the one in which the city is made inhabitable, e.g. by flooding it with water or exposing it to high levels of radioactive contamination. Yet it might be argued that, should Paris become inhabitable, it would remain Paris (at least during the first days after the change). We would not say that Paris no longer exists, but rather that Paris persists as an inhabitable city. Likewise, in case Manuel Fangio's 1956 Ferrari 290 MM just is made undrivable—e.g. by making its steering wheel stuck or extremely hot—we would not say that it no longer exists, but only that it persists as an undrivable car. On the other hand, one may parallel the contemplation of the inhabitable (and uninhabited) Paris with that of the physically intact, recently dead body of John.[2] Both may *seem* to be persisting right now. But just as we go beyond visual

[1] Not every intuition we may have will be used to *determine* whether *x* survives or not different kinds of change. For example, after inspecting its logical consequences, we may decide to drop intuition *N* because the rival intuition *N'*, whose content is that the logical consequence *L* of *N* is untenable, is stronger than *N*.

[2] As a comparative basis for exploring the Persistence Question about architectural entities and cities, we won't turn our nose up at making frequent use of the Persistence Question about persons. The comparison among architectural entities and cities, on one side, and persons, on the other side, may seem improper, if for no other reason than that a different class of items exists that appears more ontologically similar to the class of architec-

appearance in the case of John—and admit that John has ceased to exist when he died a few hours ago, no matter that his dead body still persists—similarly we may want to say that Paris has ceased to exist when it has become uninhabited a few days ago, no matter that "its dead body" is still here. Indeed, we use to speak of "dead cities" in such cases.[3] Another point in common is that both the corpse and the inhabitable city are inexorably decaying since the occurring of the event making them dead and uninhabited, respectively—so that the illusion of the persistence of Paris and John will be rapidly blown away.

A second scenario is the one in which the entire population of the city is removed and substituted with a new one, coming from a very different part of the world, speaking a different language and maybe unaware of the existence—or at least of the main characteristics—of that city until the transfer (after which, however, the city is named exactly as it was before). Suppose that we substitute the entire "Parisian" population of the actual Paris (4,366,961 persons within the "inner ring" according to the NSEE 2008 census) with the same number of

tural entities and to the class of cities—i.e. the class of ordinary objects. We may expect the metaphysical properties of a cathedral, or a city, to be more akin to those of an armchair than to yours and ours. Nonetheless we think that using persons as a comparative basis can be powerful and fruitful. One of the reasons is that, while the reasonable responses to the Persistence Question about architectural entities and cities outnumber those about ordinary objects, there is an almost one-to-one correspondence (mutatis mutandis) between the former and those about persons; and the arguments in favour and against each response are interestingly comparable. Secondly, when we care about the persistence of an ordinary object, we frequently are concerned about preserving it merely as a member of some category (e.g. the basic level category) rather than as a specific individual item. For instance, when we care about the persistence of an armchair, a refrigerator or a pair of glasses, we commonly are only interested in that they persist as members of the set of (comfortable) armchairs, (serviceable) refrigerators and (usable) glasses respectively, while the problem whether they also persist as the specific individual objects they were may easily remain out of the focus of our attention. When we deal with persons, the situation is very different: our caring about the persistence of a person is most of the times identical to our caring about the persistence of that specific individual person. Therefore, if one is interested in posing the Persistence Question about individual buildings, such as the church of Saint-Germain-des-Prés, or about a city, such as Los Angeles, a comparison with the Persistence Question about persons seems more productive.

[3] Of course, our having recourse to the Persistence Question about persons does not require cities to be persons or even organisms. Speaking of "dead cities" presupposes regarding cities as organisms, but we just take this to be a promising metaphor among many, like for example those of cities as machines, brains or political systems (Gerber and Patterson 2013; Nientied 2016). We do not agree with Varzi (2021), however, that—as robust as the analogy among cities and organisms might be—it falls in that cities do not normally "die". We would rather say that cities seem to "live" longer than any organism we know, and to withstand kinds of events that would kill any organism we know. Still we can imagine some combinations of events that would "kill" a city. Varzi writes: "Think of Hiroshima and Nagasaki. We dropped nuclear bombs on those cities. The aftermath photos are horrifying: all those buildings reduced to rubble, all those people vaporized. A devastating tragedy of incomprehensible scale. Yet the cities survived. Everything was rebuilt—homes, schools, temples, bridges, theaters." We reply that Hiroshima and Nagasaki might have survived not the nuclear bombs if, for example, all human survivors had moved to a different city, and no building was ever rebuilt. Therefore, cities *can* "die", and even *do* "die". Only, their "death conditions" are different from those of organisms.

persons coming from Shanghai. Would Paris still be Paris after the change? The question is stimulating and highly disputable. If one embraces some form of the Actor-Network Theory, for example (e.g. Lees 2001; Jenkins 2002), it will be natural to conclude that Paris is no longer Paris, since the identity of a city is considered as fixed by the complex of attitudes, experiences, intentions and emotions gravitating to (indeed *inside*) it, as realised in the minds of their inhabitants, as well as by the attributing to it of certain functions, significance, aesthetic value, and so on—all factors which cannot but dramatically vary through sets of completely different populations.[4] On the other hand, it is easy to argue in favour of the opposite conclusion. It is easy to argue, for example, that Paris has been subject to a real and full change in population from 19th century to today, and this has not even threatened its persistence through time. Of course, this population change has been continuous and gradual rather than sudden and abrupt—but why should continuous population changes lack the power to threaten the identity of cities if sudden ones do possess it? And, if we imagine to suddenly substitute the 19th century Parisian population with the present one *within the 19th century Paris*—would *this* sudden population change be lethal to Paris as well? We assume that the majority of us would doubt it would be so.

Another possible way to put an end to the existence of a city could consist in destroying, removing or saliently transforming a certain number of its most well-known landmarks and monuments. In a sense, this may be considered as an act that physically destroys some proper parts of the city. As we are exploring the ways in which we can possibly put an end to the existence of a city *other than by physically destroying it*, this kind of change may simply fall outside of our target. Yet it is intriguing to ask whether Paris would cease to be Paris should we eliminate the Eiffel Tower—we guess that *this* would not be sufficient to menace Paris' persistence; and, to ask *when* we would start to hesitate among "yes" and "no" while we add to the list (the elimination of) the Pont Neuf, the Notre Dame Cathedral, the Conciergerie, the Saint-Germain-des-Prés church, the Louvre Museum, and so on. We assume most people will agree that, whatever the point along *this* continuum at which we start being uncertain whether Paris has ceased to exist, overall *a smaller part* of Paris will have been destroyed than that that it is necessary to destroy as a whole before we start to be equally uncertain about Paris' persistence if we simply proceed by destroying one building after another from East to West, or from North to South, or by chance. Such a comparison may reveal how dependent a city's identity is from its landmarks and top tourist attractions. Interestingly, we may discover that the architectural works and spots that are most relevant for the city's persistence according to its inhabitants do not match those which are considered as the top tourist attractions.

Another fascinating scenario is the one in which a city is split into two or more new cities—or, it is merged to another city. It seems to us that it is disputable whether a city can survive these kinds of change. In particular, while someone may want to presume that, when a city *T* is split into *n* cities, one (and

[4] The reader should be aware that the supporter of the Actor-Network Theory may hold that Paris can cease to be Paris also as an effect of some change in the network of relationships lesser than a population change—such as e.g. a change in people's beliefs, desires, abilities or social status, or in their mere spatial distribution.

no more than one) in the *n* cities *must* be numerically identical to *T*, we want to deny such a presumption.

To sum up, there are at least four ways of possibly putting an end to the existence of a city other than by physically destroying it. The first way consists in making it inhabitable, and rests on the idea that a necessary condition for something to be a city is possessing a population. Thus—unlike the other three ways—making city *T* cease to exist through making it inhabitable requires making *T* cease to be a city at all. The second way consists in producing a sudden and total change in the population of the city. The idea is that a city can survive sudden *partial* changes or *slow* total changes, but not *sudden total* changes in population. If the latter occur, however, a city will continue to be a city: it will just cease to be *that* city. The third way consists in destroying, removing or saliently transforming a certain number of its landmarks and monuments. The underlying idea is that there is a critical mass of destroyed landmarks traditionally identified as distinctive of city *T* beyond which city *T* loses one of its essential properties. The fourth way consists in splitting the city into two or more cities, or, by merging it to another city. It relies on two general principles. The first principle says that, if city *T* exists at t_1 and cities *U* and *V* exist at t_2; and *U* is numerically different from *V*; and the only three possibilities are that (i) *T* at t_1 is the same city as *U* at t_2, or (ii) *T* at t_1 is the same city as *V* at t_2, or (iii) *T* has ceased to exist at t_2; and we cannot non-arbitrarily determine which of *U* and *V* at t_2 is the same city as *T* at t_1 despite knowing all the relevant facts, then (iii) is the case. The second principle says that, if cities *T* and *W* exist at t_1 and city *Z* exists at t_2; and *T* is numerically different from *W*; and the only three possibilities are that (i) *Z* at t_2 is the same city as *T* at t_1 and *W* has ceased to exist at t_2, or (ii) *Z* at t_2 is the same city as *W* at t_1 and *T* has ceased to exist at t_2, or (iii) *Z* at t_2 is a brand new city and both *T* and *W* have ceased to exist at t_2; and we cannot non-arbitrarily determine which of *T* and *W* at t_1 is the same city as *Z* at t_2 despite knowing all the relevant facts, then (iii) is the case. Note, however, that in the latter situation it is not necessary that (iii) be the case for *T* to have ceased to exist at t_2, because *T* will have ceased to exist at t_2 also if we can determine that (ii) rather than (i) is the case.

What can be said in conclusion is that, at worst, it is *open to question* whether we can put an end to the existence of a city other than by physically destroying it. Moreover, in many scenarios the intuitions and the arguments supporting a positive answer do seem no less powerful than their rivals.[5]

[5] Interestingly, one anonymous reviewer suggested that an additional way to put an end to the existence of a city other than by physically destroying it could be by *fiat*—e.g. by making it become an independent state, or several villages from an administrative point of view. We are not convinced, however, that a mere *fiat* would have the force to make a city cease to exist. Accordingly, the identity conditions over time for cities are relevantly different from those for mere institutional entities, because we can normally make an institutional entity cease to exist by simply destroying its status by a *fiat* (Jansen 2008). One could object that a specific *fiat* by the government or the safety authorities ("From this day forward, this city is off limits") *can* make a city cease to exist by making it inhabitable in the first place, no matter that no environmental condition would actually prevent it from being populated. But it seems to us that, also in such a case, the *fiat* alone is barely sufficient to make the city *inhabitable* in practice, and some supplementary physical factor is required—if only the deployment of law enforcement resources to make the ban respected. Following Weber (1921), one may suggest that a city is a space *essentially* charac-

However, there is another important scenario to be explored: relocation—the case study we want to focus on in the present paper.

3. Relocation

It is intriguing to ask what happens to a city if we relocate it, that is, if we meticulously dismantle and rebuild it in a different place on Earth, paying attention to reassemble all of its parts exactly as they were before. Imagine that no proper part of the city is physically destroyed in the operation, and that not only we use exactly the same set of physical materials—such as bricks, reinforcing steels, and so on—but also have all of them playing exactly the same roles. For the argument's clarity, suppose that also its inhabitants are equally relocated so that there is no population change—otherwise you may be observing the metaphysical effects of a population change rather than those of a mere relocation. We assume that most people will judge that no city can survive this kind of change. One may speculate that the reason resides in the resulting climate change, or perhaps in the change in the quantity and quality of the sunlight. But again—once we concede that the new location, however distant from the original one, involves no significant change for climate and sunlight—we posit that most people will maintain their opinion. It seems that relocation by itself is perceived as a serious threat to the identity of a city. The relocated item would still be a city, of course; but it would be not the same city. Suppose that we try to relocate Paris in Nevada, USA. The majority's estimation is that Paris would not survive such a relocation. But why?

We make the hypothesis that the reason is that cities do have among their essential properties that of being surrounded by a specific geographical context. Relocating a city—no matter that its population, climate and relationship with the sunlight are preserved—entails altering this essential property, hence its being lethal to the city's persistence. In other words, cities are constitutively *relational* items, and cannot survive the deprivation of their external context—i.e. the physical geographical environment surrounding them, as constituted by material entities (such as woods, hills, mountains, roads, villages, other cities, the sea) and the properties exemplified by them. Therefore, cities turn out to be metaphysically different from ordinary objects and persons, whose identity is typically untouched by relocation. Rather they are similar to geographical entities such as mountains and rivers, architectural entities such as the church of Saint-Germain-des-Prés in Paris, site-specific works of art such as *Tilted Arc* by Richard Serra (Kwon 2002; Bacchini 2017),[6] location-specific food products like

terised by the performance of some economic functions (consumption, production, and trade), so that the collapse of these functions would make the city cease to exist. Again, we doubt that this is true. We *can* imagine Paris or New Delhi to persist also in a scenario where their traditional economic functions are lost or have dramatically changed. One further interesting question is: can a city that has ceased to exist start to exist again—or, *resurrect*? If the answer is 'yes', can it do so only within a certain period of time? And, are we more inclined to acknowledge the capacity to resurrect to those cities that have ceased to exist without being physically destroyed (provided that we believe it possible for a city to cease to exist without being physically destroyed in the first place)?

[6] In 1985, Richard Serra stated that his 120-foot, Cor-Ten steel sculpture *Tilted Arc* (1981) located in Federal Plaza, New York City, was "commissioned and designed for one par-

geographical indications (Borghini 2015: 728, 735), some specific culinary works (Bacchini 2018) and—surprisingly—nano-objects, whose essential characteristics seem to depend on their environment, in virtue of the unusual ratio between bulk and surface (e.g. Bensaude-Vincent 2013).

We do not intend to deny that cities may have some other essential properties, and that some other changes different from relocation may turn out to be lethal to their persistence accordingly. But among their essential properties there is the property of being surrounded by a specific geographical context. We call this position 'contextual essentialism'. According to contextual essentialism, it is *essential* to Paris to be surrounded by the woods of Île-de-France; it is *essential* to Rome that all its ancient consular roads connect it to those quaint villages and that typical countryside; and it is *essential* to Lisbon to lie on the Tagus river estuary.[7]

According to the stronger version of contextual essentialism, it is essential to a city not simply the property of being surrounded by a specific external context, but even that of being surrounded by a specific external context *in the specific way it is surrounded by it*, where a "specific way of being surrounded by a context" is characterised, among other things, by all the spatial relations holding among the item and the context. According to the stronger version, then, a city may be threatened also by a relocation consisting in a 180-degree rotation so that the district that previously faced the sea now faces the mountains, and viceversa.

We are aware that cities' inability to survive relocation can be explained also by saying that it is essential to a city to be located exactly where it is located, that is, in the particular part of Earth's surface it occupies. This formulation may seem to pick out the same essential property we refer to, but a more careful look tells us otherwise. Indeed, you can imagine to dramatically change the context a city is surrounded by while leaving the city in the particular part of Earth's surface it occupies. On the other hand, it is equally easy to imagine moving the city away from the particular part of Earth's surface it occupies together with its context—which would apparently leave its context untouched.

Once we acknowledge that we can imaginarily manipulate either one property without affecting the other, we must of course verify the change of which property precisely is detrimental to the city's persistence. It seems to us that—if we imagine relocating Rome together with its broader geographical context (say, the whole Italian peninsula)—the relocated city would be easily judged to remain Rome. By contrast, if we envisage to leave Rome in the particular part of Earth's surface it occupies while substituting the whole Italian peninsula with—say—the Honshu island (the largest and most populous island of Japan), it is likely that the majority of people would value the transformation to be lethal to Rome. We conclude that the essential property should be correctly identified as the property of being surrounded by a specific external

ticular site: Federal Plaza. It is a site-specific work and as such not to be relocated. To remove the work is to destroy the work" (Kwon 2002: 12).

[7] The size of cities' geographical contexts can vary depending on many different factors. We will assume, however, that no city has a geographical context so small as to be negligible, and, on the other hand, that no city has a geographical context so wide as to correspond to a very large area of Earth such as, for example, a continent.

context. On the same line of reasoning, Bacchini (2017) has argued that most architectural objects (typically, buildings) are such that to change their position would be to alter one of their essential properties, where this essential property should be identified as the property of being surrounded by a specific external context, rather than the property of being located in a particular part of Earth's surface. In a sense, the present paper should be seen as an attempt to extend to cities Bacchini's view of the explanation of buildings' metaphysical vulnerability to relocation.

We do not want to deny that some people will have an intuition requiring that the essential property should be identified as the property of the city's being located in the particular part of Earth's surface it occupies. Call this position 'locative essentialism' (Casati and Varzi 2000). Basically, a locative essentialist holds that it is essential to a city to be located in the area of land it rises up in. Indeed, a locative essentialist may also want to express her position by saying that it is essential to a city to be located in a specific geographical region. But the latter formulation can also be seen as expressing the view we embrace—i.e. contextual essentialism—provided that we take a geographical region *R* to be identical with a set of features (typically specifying landforms) instantiable by one or more areas of land. If, on the contrary, we interpret a geographical region *R* to be identical with one particular area of land, then the statement according to which it is essential to a city to be located in a specific geographical region does count as a declaration of locative essentialism. But note that locative essentialism seems to be no other than a form of mereological essentialism after all (Chisholm 1973), since it can be reformulated as the view that among the essential parts of a city there are some that cannot ever be relocated—such as the particular area of a tectonic plate on Earth's lithosphere on which the city rises up, and perhaps others, like for example the "piece of sky" above it.[8] Although arguing against mereological essentialism is beyond the aims of the present paper, we just want to remark that it is a very problematic view, entailing many conclusions contrasting our common intuition (van Inwagen 2006)—especially so if applied to cities.

4. Adequate Criteria for the Persistence of Geographical Contexts

It seems to us that contextual essentialism must be coupled with adequate criteria for the persistence of geographical contexts, that is, with criteria that *do not entail* that a geographical context cannot survive any destruction or major change affecting one of its proper parts. Such combination is necessary in order to prevent a major objection, according to which the geographical context of every city we can think about—Paris, Rome, Lisbon, London—has importantly physically changed in the last centuries: villages have been created, houses have been built, forests have been destroyed, lakes have been drained, and so on. Provided that an essential property of the thirteenth century Paris is its being surrounded by its specific thirteenth century context (as we claim), positing that a geographical context cannot survive any destruction or major change affecting one of its proper parts entails that a city cannot survive it either. In other words,

[8] Nonetheless one could question the idea that the particular area of a tectonic plate on Earth's lithosphere on which it rises up, or the piece of sky above it, are *parts* of the city.

all the relevant physical changes from its thirteenth century to the present context would necessarily prove lethal to Paris. But just as Paris has survived the transformation affecting Paris itself, it has also survived the significant alteration of its context during the last centuries. So one that wants to embrace contextual essentialism must be prepared to provide criteria for the persistence of a city's context that can prevent the disastrous conclusion that a city ceases to exist as soon as just one proper part of its context changes.

One possibility is modelling such criteria on the basis of how Parfit (1984) specified the psychological criterion for the Persistence Question about persons, according to which some kind of psychological relation is a necessary and sufficient condition for a numerical identity among entities existing at different times to hold, in a case in which at least one of the entities is a person. On the psychological criterion, the correct view of the Persistence Question about persons is a reductionist view, because the fact of a person's identity over time just consists in the holding of certain more particular facts that can be described in an impersonal way and do not presuppose the identity of that person or even its existence.

The basis that Parfit takes for his own revision of the psychological criterion is Locke's view, according to which, for a thing existing in the future to count as *you* existing in the future, it is necessary and sufficient that that thing has your memories, your beliefs, your passions (although not necessarily *all* of your present memories, beliefs and passions), and some other mental states that you have now. In Parfit's terms, it is necessary and sufficient that that thing is strongly psychologically connected with you now, where psychological connectedness is the holding of particular direct psychological connections (such as, the relationships among an experience and the memory of it, or among an intention and the action that follows from it, or among a desire existing at t_1 and the same desire persisting at t_2) and *strong* psychological connectedness is the holding of very many such connections.

But the story cannot be that simple. First, Parfit adds the requirement that this psychological connectedness has not taken a "branching" form, holding between one persons and two different things. Second, as Reid first objected to Locke, identity is transitive, while psychological connectedness (whether it be strong or not) is not: I am sure that the one year old boy my parents took to Venice in 1972 is me, although I must admit that possibly no specific memory, belief or passion belonging to that boy has been inherited by me today. On Parfit's revised Lockean view, P at t_1 is the same person as Q at t_2 if and only if (i) P is psychologically continuous with Q and (ii) psychological continuity has not taken a "branching" form, where psychological continuity is defined as the psychological relation realised by overlapping chains of strong psychological connectedness. Differently from psychological connectedness, psychological continuity is transitive. While we may doubt that there are some direct memory connections between me today and the one year old boy my parents took to Venice in 1972, we can agree that there are many overlapping chains of strong psychological connectedness between them.

In analogy to Parfit's version of the psychological criterion, we may say that X at t_1 is the same city context as Y at t_2 if and only if (i) X is persistentially continuous with Y and (ii) persistential continuity has not taken a "branching" form, where persistential continuity is defined as the relation realised by overlapping chains of strong persistential connectedness; in turn, persistential connectedness is the holding of particular connections realised by unproblematic instantiations of the relationship of identity over time of entities like forests, rivers, roads, houses and villages (such as, the relationships among a river yesterday and the same river persisting today, or among a small village on Monday and the same small village persisting on Tuesday), and *strong* persistential connectedness is the holding of very many such connections.[9]

On this view, a city context can remain the same context—i.e. persist through time—also if it is affected by continuous physical transformation—just as a person persists through time in spite of her incessantly psychologically changing. The idea is that geographical contexts can persist in spite of the ongoing changing of their physical properties, regardless of whether the identity of some of the entities that are part of them is thereby distroyed. Note, however, that this is a reductionist view of the identity of contexts over time, just as is the view of personal identity over time based on the psychological criterion it is modelled after. This means that it rejects the idea that geographical contexts are separately existing entities, as well as the idea that the identity of contexts is a further fact that does not just consist in the identity of objects they are made of.

This position is able to explain why a major physical alteration of the context of a city during the last centuries (like for example that affecting the context of Paris from the thirteenth century to today) was not revealed as fatal to its persistence, even if the magnitude of the physical change may be bigger than that produced by a sudden relocation.[10]

Another basis for modelling adequate criteria for the persistence of geographical contexts may be found in Robert Nozick's closest continuer theory. According to this view, "to be something later is to be its closest continuer", where for *y* to be a continuer of *x* means that *y*'s properties are the same as *x*'s, resemble them, or at least grow out of them and are causally produced by them; for *y* to be the closest continuer of *x* means that *y* is closer to *x* than any other continuer; and closeness must be defined case by case by specifying which dimensions, or weighted sum of dimensions, determine it (Nozick 1981). Indeed, the closest continuer theory must be integrated by a theory of what closeness amounts to in the case of geographical contexts; and it is the latter theory, rather than the closest continuer theory itself, that would bear the burden of specifying the criteria of identity among contexts we look for. Thus Nozick's closest continuer theory seems to be more a complement to a

[9] Indeed, Parfit distinguishes among a narrow view, which also requires that psychological continuity have the right kind of cause, and two wide versions, that allow any reliable cause, or any cause, respectively. The same distinction can be drawn with regard to persistential continuity.

[10] We are aware it could be questioned that psychological continuity is a necessary and sufficient condition for personal identity over time. Firstly, it may not be a necessary condition since apparently a temporary mental blackout briefly shutting down all psychological connections would not be detrimental to the persistence of a person if followed by a restart of mental life in the very same configuration it possessed before. Secondly, it may not be a sufficient condition since some slow yet severe and irreversible kinds of psychological transformation (say, gradual and permanent demonic possession) may count as lethal to personal identity in spite of their being compatible with the holding of overlapping chains of strong psychological connectedness. Mutatis mutandis, the same worries could be raised with regards to persistential continuity as a necessary and sufficient condition for a city context identity over time.

view of continuity under identity like that outlined above than one of its rivals. Furthermore note that, as conceded by Nozick in general, a context may be the closest continuer to the context of city *T* without being close enough to it to be the context of city *T*. In other words, being the closest continuer of a certain context is at best only a necessary condition for being identical to that context.

5. Can a City Ever Survive Relocation?

One of the consequences of contextual essentialism is that *some relocations* of a city may not alter the city's identity, provided that also the context is relocated (and, its identity survives the change). However, contextual essentialism is clearly also compatible with the fact that no city relocation is ever possible; in fact, it might turn out that no city context can ever be relocated.

Consider that, on contextual essentialism, it is also possible that cities can survive some kinds of relocation also if cities contexts cannot ever survive any relocation. This is possible, for example, if we conceive geographical contexts as regions of space rather than complex (spatial) relations nets characterising single spots. In this situation, replacing a city *inside its original geographical context*, also if in a different position within it, would count as relocating it while preserving its context. In any case, as long as city contexts can be relocated, ceteris paribus also cities can be relocated.

City contexts might turn out to be immovable for a number of different reasons. For one thing, it might be that (differently from cities) city contexts essentially hold the property of being located in the particular part of Earth's surface they occupy. Or, suppose that they—just like cities—do have among their essential properties that of being surrounded by a specific *broader* geographical context. The position according to which a geographical context could only be relocated by relocating *its broader geographical context* may seem affected by an infinite regression; as a consequence, nothing could ever survive relocation that has some geographical contextual properties among its essential properties in the first place.

We believe that the infinite regression problem can be solved. Note that the solution we provide allows holding that any geographical entity or region of space—regardless of how extended it is—has among its essential properties that of being surrounded by a specific broader geographical context. Consider first an architectural entity like a building. Suppose you maintain that among its essential properties there is the property of being surrounded by a specific material context; and, call this context the "urban context" of the building (supposing that the building rises up in a city). We want now to distinguish between the city the building rises up in, on one hand, and the building's urban context, on the other hand. These are two different items admitting of different persistence conditions. In particular, relocating the urban context seems to us easier than relocating the city. In order to make the urban context survive relocation, you may only need to preserve its physical identity or even physical continuity. In other words, the urban context—differently from the city—does not seem to have among its essential properties that of being surrounded by a specific broader geographical context. So the church of Saint-Germain-des-Prés can only survive relocation if its "Parisian context" is preserved; but it is possible to hold that the persistence of this "Parisian context" tolerates relocation much better than Paris itself, so that relocating the church of Saint-

Germain-des-Prés is not affected by the difficulty of relocating Paris in the first place.

The same line of reasoning holds for larger items such as cities. Like the church of Saint-Germain-des-Prés, Paris is an essentially context-dependent item. But when we distinguish among Paris' geographical context, on one side, and the region of Île-de-France, on the other side, we are able to posit that only the latter is in turn characterised by having among its essential properties that of being surrounded by a specific *broader* geographical context. Then we can envisage moving the geographical context Paris is essentially dependent on without necessarily moving the region of Île-de-France. This makes Paris movable in spite of both Paris and the region of Île-de-France having among their essential properties the property of being surrounded by their own specific geographical context. There is no infinite regression.

Regardless of whether we want to distinguish among a geographical region and the geographical context of an item (like a city) rising up in that region, of course, it is still possible that the infinite regression holds if also geographical contexts—like geographical regions—are revealed as being essentially context-dependent. Moreover, also in case they are not so, and accordingly there is no infinite regression, city contexts might turn out to be incapable to survive relocation because of some other reason.

One possibility is that a city context is immovable because of its being *vague*. If it is indeterminate whether one or more areas belong to the context, it might be impossible to determine where it exactly lies and hence what exactly has to be relocated. If the context's boundaries can be fixed only arbitrarily, then it seems impossible to decide which of an infinite list of partially overlapping geographical contexts should be moved. To make matters worse, vagueness involves a pernicious puzzle, i.e. the sorites paradox. In fact, assuming that an area A_1 belongs to the city context C, arguably an area A_2 adjacent to A_1 belongs to C too. By induction, any area A_n belongs to C, included any area that may lie thousands of miles away from C.

How can we solve this problem? Varzi (2001), following Russell (1923) and Lewis (1986), argues that vagueness in the geographical domain is semantic, not ontological. Namely vagueness is a feature of the terms by means of which we pick out geographical objects, rather than being a feature of the objects themselves. Thus, we can get around the problem by adopting an adequate semantic approach, like for example supervalutationism. The basic idea under a supervalutationistic semantics is that the name 'context of city *T*' is vague—i.e. there are some specific portions of Earth's surfaces that neither determinately are nor determinately are not the context of city *T*, or equivalently, there are some areas that neither determinately belong nor determinately belong not to the context of city *T*—because the name 'context of city *T*' admits of many different legitimate referents. When we put in making the meaning of a vague name or predicate more precise, we accordingly have many legitimate ways of doing it. Each way of making a vague name or predicate more precise is a *precisification*. A precisification is admissible if and only if every sentence that is determinately true (false) in English is true (false) in the precisification (Weatherson 2016).

Consider the statement B = "X belongs to the context of city T". Call S the set of all the areas (or even parcels of land) A_1, A_2, …., A_n such that substituting X with each area A_i, B is true under every admissible precisification of the predicate 'belonging to the context of city *T*' (or, equivalently, of the name 'context of city *T*'). We call S the 'minimal context of city *T*'. Then call *S′* the set of all areas A'_1, A'_2, …., A'_n such that substituting *X* with each area A'_i, *B* is true under some admissible precisifications of the predicate 'belonging to the context of city *T*', and false under others. We call 'enlarged context of city *T*' every area mereologically composed by both *S* and at least one area that is a member of *S′*. We call the 'maximal context of city *T*' the biggest of the enlarged contexts of city *T*, that is, the one enlarged context of city *T* that includes all the members of *S′* as its proper parts. If you want to adopt a strict view of how vagueness must be contrasted, then what has to be moved in order to move the context of city *T* is the minimal context of city *T*. If you want to adopt a more liberal view, however, you can move any of the enlarged contexts of city *T*, included its maximal context. In any case vagueness is no longer a problem. In order to be able to relocate a city context, at worst you might have to previously pick it out from a set of equally good candidates—that is, just in case you adopt the liberal rather than the strict view. Note that each enlarged context seems to fully possess the status of being the geographical context of that particular city under the liberal view; and that apparently there is no difficulty for a city to have more than one geographical context.

Vagueness may also make it arbitrary to distinguish between the city and its context in the first place. Also this difficulty—however less serious to contextual essentialism—can be treated using the same approach; first we can identify as city *T* the minimal city *T* or else any of the enlarged cities *T*, and then we can identify its context as specified above.

6. Conclusions

There is at least one notable difference among the metaphysical nature of cities, on one side, and that of ordinary objects and persons, on the other side. The identity of ordinary objects and persons over time is normally thought to be untouched by variations in location. Ordinary objects like chairs, apples and books can be moved without threatening their identity. Similarly, people are normally considered to be the same after they have travelled or when they move to another country, and we ordinarily accept that anyone can survive her permanently moving from Paris to Tokyo if no particular accident occurs. By contrast, cities are not thought to normally survive relocation. Like architectural and geographical entities, site-specific works of art, location-specific food products, specific culinary works and nano-objects, cities seem to be very vulnerable to relocation.

We have advanced a view accounting for this fact, called 'contextual essentialism', according to which cities do have among their essential properties the extrinsic property of being surrounded by a specific geographical context. Cities turn out to be essentially relational, context-dependent items. We have shown how contextual essentialism is a better account of the metaphysics of cities than its main rival, i.e. locative essentialism. We have concluded that a necessary condition for a city to persist over time is the persisting over time of its context; and we have outlined a view of a city context's identity which is capable of explaining why some major physical alteration of the context of a city, such as that affecting the context of Paris from the thirteenth century to today, was not revealed as fatal to its persistence, even if the magnitude of the physical change is probably bigger than that produced by a sudden relocation.

If we are right, a city could be relocated in principle, since—as we have shown—there might be no metaphysical obstacle to moving a city context. In fact, geographical contexts—as distinguished from geographical regions—may not be essentially context-dependent in turn; and the difficulties normally due to vagueness in the geographical domain can be solved by adopting a specific view of what vagueness is as well as an adequate semantic approach in order to dispel its fog.

We are aware that essentialism is not particularly trendy today in metaphysics in any of its versions. We should be prepared, however, to acknowledge essential properties whenever the explanatory advantages exceed the costs.[11]

References

Bacchini, F. 2017, "The Persistence of Buildings and the Context Problem", in K. August and L. Schrijver (eds.), *Analytic Philosophy and Architecture. Approaching Things from the Other Side*, Delft: Japsam Books, 85-104.

Bacchini, F. 2018, "The Ontology of the Architectural Work and Its Closeness to the Culinary Work", *City, Territory and Architecture*, 5, 1, 5-21.

Bensaude-Vincent, B. 2013, "Decentring Nanoethics toward Objects", *Ethics & Politics*, 15, 1, 310-20.

Borghini, A. 2015, "What Is a Recipe?", *Journal of Agricultural and Environmental Ethics*, 28, 4, 719-38.

Casati, R. and A.C. Varzi 2000, "Topological Essentialism", *Philosophical Studies*, 100, 3, 217-36.

Chisholm, R.M. 1973, "Parts as Essential to Their Wholes", *Review of Metaphysics*, 26, 581-603.

Gerber, A. and B. Patterson (eds.) 2013, *Metaphors in Architecture and Urbanism*, Bielefeld: transcript Verlag.

Lees, L. 2001, "Towards a Critical Geography of Architecture: The Case of an Ersatz Colosseum", *Ecumene*, 8, 1, 51-86.

Jansen, L. 2008, "The Diachronic Identity of Social Entities", in C. Kanzian (ed.), *Persistence*, Heusenstamm: Ontos Verlag, 49-72.

Jenkins, L. 2002, "Geography and Architecture. 11, Rue du Conservatoire and the Permeability of Buildings", *Space and Culture*, 5, 3, 222-36.

Kwon, M. 2002, *One Place After Another. Site-specific Art and Locational Identity*, Cambridge, MA: MIT Press.

Lewis, D. 1986, *On the Plurality of Worlds*, Oxford: Blackwell.

[11] We would like to thank two anonymous reviewers for their constructive comments and criticism. Also, we would like to thank Massimo Dell'Utri, Donatella Donati, Andrea Borghini, Giuliano Torrengo and all the participants in the 2018 conference *On the (Im)Possible Futures of Cities and Agricultural Sites*, The Center for Philosophy of Time, University of Milan, for their comments on earlier versions of this paper.

Nientied, P. 2016, "Metaphor and Urban Studies-A Crossover, Theory and a Case Study of SS Rotterdam", *City, Territory and Architecture*, 3, 21, 1-10.

Nozick, R. 1981, *Philosophical Explanations*, Cambridge, MA: Belknap Press.

Parfit, D. 1984, *Reasons and Persons*, Oxford: Oxford University Press.

Russell, B. 1923, "Vagueness", *Australasian Journal of Psychology and Philosophy*, 1, 84-92.

van Inwagen, P. 2006, "Can Mereological Sums Change Their Parts?", *Journal of Philosophy*, 103, 12, 614-30.

Varzi, A.C. 2001, "Vagueness in Geography", *Philosophy and Geography*, 4, 1, 49-65.

Varzi, A.C. 2021, "What is a City?", *Topoi*, 40, 2, 399-408.

Weatherson, B. 2016, "The Problem of the Many", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2016 Edition.

Weber, M. 1921, *The City*, New York: Free Press.

Williamson, T. 1994, *Vagueness*, London: Routledge.

What Galileo Said

John Biro

University of Florida

Certainly there are unchanging truths,
but there are changing truths also,
and it is a pity if logic ignores these.
(Arthur Prior)

Abstract

Davidson's paratactic account of indirect speech has it that a natural-language report of an utterance such as Galileo's supposed one of 'The Earth moves' should be understood as analyzable into two separate, and semantically independent, utterances, the first of which points to the second, with the latter meaning in the reporter's mouth what Galileo's meant in his. The account rests on the assumption—shared by most writers on the subject, including critics of the account—that the correct natural-language report of Galileo's utterance is 'Galileo said that the Earth moves.' I show that on that assumption the paratactic analysis misfires: the two utterances—Galileo's and the reporter's—do not samesay one another. However, this is also the case if the verb in the demonstrated sentence is changed to respect the tense-sequencing rule as does 'Galileo said that the Earth moved.' Since the latter does correctly report Galileo, that must be because, contrary to the central claim of the paratactic analysis, its two clauses are not semantically independent.

Keywords: Indirect speech, Donald Davidson, Parataxis, Tense sequencing, Misreporting.

That there are difficulties in accounting for binding of various sorts across clauses in indirect discourse is has been known for a long time. (Higginbotham 1986, *inter alia*) Some of these have been thought to present problems for the paratactic account, given its central thesis that a report of the form '*S* said that *p* ' is, contrary to appearance, not really an utterance of a single sentence but of two separate and semantically independent sentences, the utterance of the first by the reporter asserting that x said what the utterance of the second by the reporter says. For the report to be true, the reporter's second utterance must "samesay"—that is, have the same content as—*S*'s utterance. Yet while two sentences

ISSN 2465-2334

DOI 10.14275/2465-2334/20200.bir

containing different but co-referring expressions—names, pronouns, definite descriptions—can express the same proposition, they are not substitutable *salva veritate* in indirect reports (any more than in other opaque contexts). Which of these should we choose as the vehicle for specifying the content of *S*'s utterance?

Here I shall not be concerned with the so-called paratactic gap that opens up as a result nor with attempts to bridge it (e.g., Blair). Instead, I shall adduce some reasons for thinking that there is another kind of gap, having to do with tense, that cannot be closed.

In summarizing the paratactic account, Burge notes in passing that it requires ignoring tense (1986: 192). This, I shall show, results in mis-reporting the speaker. To get our report right we have to take tense into account, and we can do so through the machinery of tense sequencing. However, a paratactic report is unable to accommodate tense in that way. This stands in the way of securing the samesaying relation central to the account.

Suppose Galileo to have uttered

(1) The Earth moves.[1]

It is assumed by Davidson that the proper natural-language report of what he said is

(2) Galileo said that the Earth moves.[2]

Davidson's so-called paratactic analysis of (2) is as two separate and semantically unrelated sentences, as in

(3) Galileo said that. The Earth moves.

In (3), 'that' is to be seen as a demonstrative, pointing to an utterance of its second sentence.[3] What Galileo is said to have said is what the reporter would be saying were he to utter that sentence: the two speakers (or the two utterances) are said to *samesay* each other. For this to be the case, it is not necessary that the utterances they respectively make be of synonymous sentences. Samesaying is a relation not between sentences but utterances. As Davidson—somewhat misleadingly—puts it, because "Galileo uttered a sentence that meant in his mouth what 'The earth moves' means now in mine" (1968: 140). Two utterances of sentences with different meanings can samesay each other, as with my utterance

[1] What he is supposed in legend to have uttered is, of course, not that sentence but 'Eppur si muove.' But we can go along with Davidson in pretending. However, 'The Earth' is a proper name, the capital 'T' being part of it. Davidson's (and others') 'the Earth' is a mutation. (True, these days the definite article is sometimes dropped, bringing our planet's name in line with fellow planets. We do not say: 'The Saturn.' But then it should be dropped from the content clause of the paratactic report, as well.)

[2] That this is the proper report usually goes unquestioned. McDowell, Higginbotham, Larson and Ludlow and Ludwig and Ray all assume it. The last-named list nine difficulties for the paratactic account that they claim their rival (though still, in spirit, Davidsonian) account solves. The problem I shall suggest arises from accepting (2) as the proper natural-language report and fashioning a paratactic version based on it is not among them.

[3] For present purposes, I shall grant Davidson's claim that 'that' in (2) is a demonstrative pronoun. Misgivings about that claim may be found in Segal and Speas 1986, Segal 1989, and Biro 2011. But, as Blair argues, some version of the paratactic account may work even Davidson is wrong about this.

of 'I am tired' and yours of 'He is tired' or with an utterance of 'We visited the capital of Italy' and one of 'We visited the Eternal City'.[4]

Suppose, then, Galileo to have uttered a token of the sentence 'Eppur si muove.' As Higginbotham (who also accepts (2) as a correct report) points out, Italian unlike English "permit[s] a simple present-tense, non-progressive sentence to be a report on the current scene" (1999: 213). Let us take Galileo to have intended, as we know he did not, such a report, wanting to say (only) that the Earth was moving at the time of his utterance. What would be the correct paratactic report of what he said? It cannot be

(4) Galileo said that. The Earth is moving.

for the second sentence of (4) in my mouth means that the Earth is moving at the time of my utterance. While Galileo may well have believed that this would be the case, his uttering what he did is not sufficient for reporting him as saying that it would be. One may say that the Earth is moving without believing that it would be doing so at some time later. Nor can it be

(5) Galileo said that. The Earth was moving.

for he did not utter anything that in his mouth meant what its second sentence means in mine. And if he had, he would have been saying that the Earth had been moving at a time earlier than his utterance and implying that it was no longer doing so.[5]

Higginbotham maintains that 'Mary said that a unicorn was walking' may be a report of an utterance by Mary of either 'A unicorn is walking' or 'A unicorn was walking' (1999: 200). Not only is this hard to square with his official doctrine that sequence of tense is obligatory (2009: Ch. V), but it seems plainly wrong: had Mary uttered the latter sentence, we would surely report her as having said that a unicorn *had been walking*. Nor is this a slip: we find the same claim, that an indirect report whose content clause is in the simple past may be a correct report of an utterance itself in the simple past, in his 2009 (83-84), where he suggests that an utterance of 'Gianni said that Maria was ill' could be made true by Gianni's saying, sometime in the past, "something to the effect" that Maria was ill *at the time*. That something would be, presumably, his saying (uttering) 'Maria is ill.' Yet, according to Higginbotham, that same report can also "constitute a report of a past past-oriented utterance." This cannot be right: if Gianni had wanted to say something about Maria's health prior to his utterance, he would have uttered 'Maria was ill' and our report of what he said would have

[4] Some authors seem to assume that samesaying requires synonymy (Elugardo 1999, Burge 1986). But see Davidson 1999. Davidson himself speaks of two utterances' having the same content as their "translating" one another (1976).

[5] An utterance of (2) could also mean something quite different, as could that of 'Galileo said that the earth was moving': the former that the planet was on its way to a different place in the firmament, the latter that an earthquake was in progress. Translating Galileo's 'si muove' either as 'turns' or as 'revolves', rather than 'moves,' avoids these obviously irrelevant interpretations, making it clear that we mean either that the Earth spins around its axis or that it orbits the sun. (Many languages mark the difference between these lexically: German has 'dreht sich' and 'umkreist' for 'spins' and 'orbits,' respectively, and 'bewegt sich' for 'moves', Hungarian 'forog' (spins) and 'kering' (orbits)—and 'mozog' or '(meg)mozdul' for 'is moving' and 'moves,' respectively.)

to be 'Gianni said that Maria had been ill.' Gianni cannot say that Maria is ill at the time of his speaking by saying 'Maria was ill,' and he cannot be reported as having said this by 'Gianni said that Maria was ill.' In the same way, if we were to take (4) as saying that Galileo uttered what follows 'that' and (5) as saying that he uttered its second sentence, we would be representing him as having said that the Earth had moved but had come to rest prior to his speaking. Obviously, that would be getting him wrong, but what else can we take (4) and (5) to be saying?

Thus Davidson is wrong when he says: "Galileo utters his words 'Eppur si muove,' I utter my words, 'The Earth moves.' There is no problem yet in recognizing that we are samesayers; an utterance of mine matches an utterance of his in purport. I am not now using my words to refer to a sentence; I speak for myself, and my words refer in the usual way to The Earth and to its movement" (1968: 141).

If 'moves' in the second clause of (2) and in the second sentence of (3) cannot be interpreted as 'is moving,' can it be interpreted as being about the time of (2) and (3) are uttered, as the paratactic account requires? Perhaps we can make Galileo and me samesayers, after all, by understanding (2) and (3) as saying that we both said that the Earth moves *now*, where 'now' is indexed to the time of our respective utterances. However, the familiar distinction between meaning and reference (or, as some say, character and content) must be kept in mind here. While what the sentence Galileo and I both utter has but one meaning, 'now' (or 'at the time of speaking') has a different reference in our respective utterances of it. Given this, we assert different propositions by uttering the same sentence and are thus not *saying* the same thing in any recognizable sense. Thus if we take 'si muove' to mean either 'is moving' or 'moves now,' (3) cannot be an accurate report of what Galileo said. In different ways, both these interpretations make the report into one "on the current scene."

In an everyday report that respects tense sequencing we have no difficulty in avoiding such misrepresentation.[6] We say

(6) Galileo said that the Earth moved.

This cancels the unwanted implication, present in both (2) and (3), that when uttering what he did, Galileo was saying something about what the Earth would be doing in 2019. This is desirable, since while it is a reasonable assumption that he, and anyone else, uttering the sentence in 1633 would have been disposed to assert that it would be, the fact that he uttered what he did is not enough for us

[6] The tense-sequencing rule for indirect discourse is (roughly speaking) that the tense of the verb in the content clause of an indirect must shift to the past perfect, the pluperfect, or the conditional, respectively, according to whether its tense in the original utterance is present, past, or future. Thus 'It is raining' goes to 'He said that it was raining,' 'It was raining' to 'He said that it was raining,' and 'It will rain' to 'He said that it would rain.' Not all languages have such a 'backshifting' or 'attracted sequence' rule. Speakers of languages that do not have to rely on context and collateral information to interpret a report such as (2). And it is sometimes thought that even in languages that do have such a rule, such as English, it admits of exceptions. I discuss the example Lepore and Ludwig offer (2003: 98-99) and their gloss on it below.

to say that he did so.[7] We must leave it open that were he with us today, something between then and now might have changed his mind.[8]

Can the paratactic account do this? If it respected the tense-sequencing rule, the report would come out, presumably, as

(7) Galileo said that. The Earth moved.

It may be objected that the Italian sentence Galileo uttered was in the present tense, hence it and the second sentence in (5) differ in meaning and, if so, the latter, as uttered by the reporter, cannot samesay Galileo's utterance. But, as already noted, two sentences need not be synonymous for the utterances made using them to samesay each other. The trouble with (5) lies elsewhere.

If we make the paratactic reformulation of the report respect the tense-sequencing rule, as in (7), it is hard to accommodate the demonstrative aspect of the account. Just what is the referent of 'that' in (7)? There is nothing in the offing but the second sentence. Of course, he need not—would not—have uttered that. Nor, as noted earlier, need he have uttered a sentence synonymous with it. But the paratactic account requires that he uttered something that samesays an utterance of that sentence in my mouth, and it is hard to see what that could have been. Had he uttered the demonstrated sentence of (7), he would have said that the Earth had moved at some time before he spoke. What he uttered was (we are supposing), Italian for 'The Earth moves' (understood as a report on the present scene). But, as we have seen, taking the demonstrative to refer to that sentence represents him as having said what I would be saying in uttering that sentence. As Davidson says in dismissing Fregean approaches that posit a difference of sense in direct and in indirect contexts for the same expression, that it is "plainly incredible that the words 'The earth moves', uttered after the words "Galileo said that", [should] mean anything different, or refer to anything else, than is their wont when they come in other environments" (1968: 144). Yet, surely, if I uttered the sentence in the second sentence of (4) by itself, I would not be saying something quite different from what Galileo can be supposed to

[7] If I hear someone utter 'The Earth moves', I do interpret him, other things being equal, as saying that it has been, and would continue to be, in motion. But while to use the simple present tense is (often, though not always) to suggest that the state in question is continuous, it is not to assert this, as is shown by the ease with which the suggestion may be cancelled. ('The Earth moves but may stop doing so if ...') Conversely, using the continuous present suggests, but does not entail, that the state is a merely temporary one. ('The Earth is moving and will continue to do so.') And languages that make no distinction between the simple present and the continuous present (or past), such as German and Hungarian, still require tense sequencing.

[8] It is even possible, for all his uttering the sentence in 1633 entails, that even then he did not believe that The Earth would continue to move after his utterance. (See also fns. 11 and 12 below.) Unlikely, of course, and we have good non-semantic reason to interpret him as believing that it would. I am assuming here that (4) is a warranted interpretation of what Galileo uttered. Indirect reports always involve interpretation, and the ways in which that and the attribution of belief based on it proceeds and the pitfalls it involves raise subtle and complicated questions. Some of these are discussed in Biro (1984, 1992). Here what is in question is only whether an otherwise warranted interpretation could have the underlying form the paratactic account says it does.

have said. We would both be saying that the Earth was moving at the time of our utterance.

The argument thus far has been intended to show that the paratactic account cannot be a general account of indirect reports, as it cannot capture those made with a verb in the continuous present. Does it even work for utterances with a verb in the simple present? Can we, by interpreting Galileo's 'Eppur si muove' as being *not* about the current scene, as it was obviously not intended to be, secure the samesaying relation between his utterance and the second sentence of (3)?

We can take Galileo to be saying something about whether the Earth would be moving now, when *we* speak (as well, of course, when he spoke), if we understand the verb in the simple present to be used in the habitual sense, as when we say of someone that he smokes or goes to church. This is the line urged by Lepore and Ludwig (2003), who claim that the acceptability (as they think) of

(8) The Egyptians knew that the Earth is round.

shows that "reports of certain states that continue into the present" are exceptions to the strict tense-sequencing rule (2003: 98).[9] They suggest that this is so with indirect reports, too. If so, (2) is also acceptable, for the same reason. In such cases, they say, "if we wish the reports to be possibly true, the right account should focus on what the reportee knows, hears, says, and the like." If this is sound advice, as I think it is, following it invites us to respect the tense-sequencing rule, rather than flouting it, as (2) and (3) do. If we take 'is round' in analogues of these to mean what it means in an analogue of (1), (8) is false. The Egyptians knew no such thing. Assuming that we have satisfied ourselves that they had done their work, we are justified in saying that they knew that the Earth *was* round, but *only* that, certainly not that they knew what the shape of the Earth would be in 2019. Of course,

(9) The Egyptians believed that the Earth was, and would continue to be, round.

may well be true (and we may allow that they may have been justified in their belief). And perhaps

(10) The Egyptians believed that the Earth is round.

[9] Compare Larson and Ludlow, who say that "if one wished to report in English what a speaker of German said in uttering 'Galileo glaubte dass die Erde sich bewegte' it would be very natural to employ 'Galileo believed the Earth moves.'" (1993: 334). They do not even note the change from the past-tense 'bewegte' to the present-tense 'moves' and take the latter to samesay the former. While they do not say so, one can conjecture that the reason for this is the same as the one suggested by Lepore and Ludwig. The latter "suspect that in this case the present tense is used to indicate that the content of the reported state or event is not relativized simply to the time of the reported event or state, but is about a state that would extend from that past time into some indefinite future time that at least includes the time of utterance" (98). However, Lepore and Ludwig also note that such an exception is possible only if the main verb is factive. They give as an example of where it is not 'I thought that the Earth is round.' If so, 'The Egyptians believed that the Earth is round' should be ruled out, as well, as should Larson and Ludlow's translation of the German. (Perhaps it is the fact that one is reporting on oneself that makes the example Lepore and Ludwig give unacceptable.)

can be interpreted to mean what (9) does. However, to do so we need more to go on than the Egyptians' uttering

(11) The Earth is round.

Their doing so falls far short of being evidence for that. The fact that the past tense in the main clause of (9) requires the subjunctive in the second conjunct of its embedded clause is an indication of this.

In the same way, we should not say of Galileo, great scientist that he was, that he knew that the Earth moves (habitual) now, only that he knew that the Earth moved (habitual) then. Even if he believed that the Earth would continue to do so, he would not have claimed to know—that is, to be certain—that it would. Neither Galileo nor the Egyptians claimed to be soothsayers, and they should not be represented as such. But that is what we are doing with (2) and (8). Thus the fact that 'moves' in (2) *permits* the habit to be a continuing one accommodates the fact that Galileo presumably intended to express by it something to the effect that he believed that, in the nature of things, the Earth was, and always would be, in motion. But even the nature of things can change, and we should not report Galileo as asserting the contrary, even if we are convinced, and no doubt rightly, that he believed that it could not. The information on which that conviction is based is not semantic, and the machinery of indirect discourse should operate in the same way whether the reporter has such information or not. In particular, it should treat cases in which it is plausible to interpret the verb in the target utterance as habitual and cases where it is clearly not.

Compare these three reports:

(12) I had a call from our friend. He said that he was in Paris.

(13) I had a call from our friend. He said that he is in Paris.

(14) I just had a call from our friend. He said that he is in Paris.

Assume that what the friend uttered was 'I am in Paris.' The difference between (12) and (13), on the one hand, and (14) on the other, is that the first two leave unspecified how long before the report the friend spoke, whereas the third tells us that it was very recently. Suppose we are reporting on yesterday's call. (12) would be clearly true, but (13) could not be. It says, falsely, that the friend said that he is in Paris today. (It would be true if the friend had said 'I will be in Paris tomorrow.') Unless the friend added to 'I am in Paris' something like 'and am staying for a day or two,' I am not entitled to report him with (13). (This is even clearer if we imagine him adding 'and will be in London tomorrow'.) In (14), the first sentence indicates the proximity of the report to the utterance reported on, which ensures that, absent funny business, the report is true. But the information that ensures this is contained in the first sentence, which is one uttered by the reporter, not the reportee. There is nothing in what the *latter* uttered that guarantees the truth of (14).

Of course, here, too, I may have information that the friend would be staying on in Paris independently of what he said during the call which would make it reasonable for me to utter (13). But that does not make it the correct report of what the friend said.

The reason why (14)'s first sentence is naturally followed by one with a present-tense verb is not just that it is unlikely that he would have left Paris in the short interval between his utterance and the reporter's. Even when that interval is long enough, and we know that it is, it can be natural to report in the present tense. I receive a letter from our friend that begins ‘I am in Paris,’ and I say to you ‘Our friend says (that) he is in Paris.’ We both know that he may well have moved on since he wrote the letter, but the tense of the verbs reflects our understanding that it refers to the time of writing not the time of reading. In saying that our friend says that he is in Paris, I am not reporting in the sense in which I am with (12) but am merely repeating his utterance (with the pronoun changed, of course). That the present-tense verb refers to the time of my utterance is conveyed by the first sentence of (14). The present-tense verb in the second sentence does not by itself refer to any determinate time. By contrast, the past-tense verb in the second sentence of (12) tells us that the report (if not necessarily the call) is subsequent to the present-tense utterance (‘I am in Paris’) it reports.

The paratactic analysis offers no way to capture these facts. Neither ‘He said that. I am in Paris’ nor “He said that. I was in Paris’ say what the second sentence of (12) says. And, as just suggested, the second sentence of (14) is not really an indirect report in the way (12) is but a case of passing on an utterance in itself semantically un-interpreted, as we do in direct quotation, along with a (pragmatic) pointer in the first sentence as to how to interpret it.

If we adopted the paratactic model for reports of what someone knew, as in

(15) The Egyptians knew that. The Earth is round.

we would be, on a natural interpretation, getting them wrong in the same way as we get Galileo wrong with (3). Interpreting the present-tense verb in the second sentence as relative to the speaker’s context guarantees this. The fact that *we* know that the Earth is still round does not entitle us to attribute knowledge to the Egyptians that it would be in 2019. What matters is not whether we think that the state involved in the report is one that has continued to this day but whether we are justified in interpreting the reportee as knowing that it would. However, his uttering something in the present tense falls short as evidence for so interpreting him. Again, I am not saying that we may not have good reason to attribute the beliefs these reports do to Galileo and the Egyptians, respectively. But that we have such reasons, when we do, is not a fact about the semantics of indirect reports. Neither our natural-language reports nor a theory about their underlying semantic structure should suggest otherwise. We should no more accept (2) than its paratactic offspring.

But how else to interpret (8)? Here, again, we face the same dilemma as we did with (3). We are told to interpret the content-sentence relative to the reportee’s context, which means that we cannot take the present-tense verb to be making a claim about the Egyptians’ knowledge of the future habits of the Earth. Yet, as Lepore and Ludwig insist with respect to (3), that is what we would need to do to make (8) and its paratactic version (15) true. At the same time, we are asked to interpret the former’s content clause and the second sentence of the latter as meaning what they would mean in other contexts, including one in which they are uttered by themselves. Not only does Davidson, too, insist on this, as we have seen—we really cannot help doing so. These two injunctions pull in different directions.

Perhaps we can avoid the dilemma if we get our everyday report right, as we would be doing with

(16) The Egyptians knew that the Earth was round.

This has as its paratactic re-formulation, presumably,

(17) The Egyptians knew that. The Earth was round.

Our problem is still with us, though, as it was with (2). The Egyptians would not have used the embedded sentence of (16) to say what (we want to say) they knew, for in their mouth, that sentence would have meant what we can say only by saying 'The Earth had been round'. Still, it may be said, *we* can use it to do so. But, once again, the sentence being demonstrated is one whose natural interpretation is one relative to the reporter's context. That, in fact, is essential to the report's getting it right: the Egyptians' uttering the demonstrated sentence tells us that they believed that something had been the case and *only* that. The sentence that the Egyptians would have uttered if they wanted to say that they believed that the Earth was round at the time of their uttering is not the second sentence of (17) but (15).

It should be noted that 'said' denotes an action, not a state, as do 'believe,' 'know,' and the like. Lepore and Ludwig treat these as on a par. However, even if we accepted their claims about reports involving the factive states such as knowing, as I have argued we should not, the dilemma the paratactic account of indirect discourse runs into remains, and it can be put in a nutshell. The content-sentences of (2) and (3) and the sentence Galileo uttered mean the same thing, but the utterances made by uttering them do not samesay each other. On the other hand, the content-sentence of (6), which is, as I have urged, the correct natural-language way to report Galileo, is not something (whose Italian translation) Galileo ever uttered. More importantly, had he done so, he would have said something different from what he in fact said. Thus in pointing to it, we would be pointing to the wrong thing. Galileo and I cannot say the same thing by uttering (1) any more than you and I can say the same thing by uttering 'I am hungry.'[10] But, unlike in the case of pronouns and other indexicals, with tensed verbs we cannot say the same thing with different sentences, either.[11] Not only do the content clause of (6) and the second sentence of (3) differ in meaning, the utterances Galileo and I can make if we use them express different propositions. As we have seen, the utterance attributed to him by (2) would have been true if and only if he had said that the Earth was moving in 1633. The report I would make if I used (2) would be true if and only if Galileo had said that the Earth would be moving in 2019.

A last-ditch attempt to save the paratactic account may take the following line. Taking a hint from Lepore and Ludwig, we may argue that (2) is, after all, acceptable as a report of what Galileo said, as the property it has Galileo attributing to the Earth is one it is plausible to think he thought it would continue to possess, hence it is plausible to think that he said it would. If this is so, our account of the semantics of his utterance, if not that of the semantics of his sentence, should be sensitive to this. But this will not do for two reasons. First, it

[10] If not, that is not because of the difference in pronouns. Examples that do no involve such a difference abound: 'Churchill and I cannot say the same thing with 'Germany is a menace to civilization,' nor Babe Ruth and I with 'The series is fixed' or Galileo with 'The Inquisition is powerful.' 'He says that he will vote Tory' and 'He said that he would vote Tory' both make sense—but does 'He said that he will vote Tory'?

[11] Arguably, there is a sense of 'say' in which 'I am hungry,' said by me and 'You are hungry,' said by you to me do not really *say* the same thing. But we can allow that there is a sense in which we express the same proposition, which is enough for present purposes.

would make it impossible to interpret someone uttering what Galileo did as having said that the Earth moved at a particular time and at that time only. Second, even if we allowed that this was (2)'s correct, or, at least, default, interpretation in this special class of cases, this would not help us to a general account of indirect discourse, which the paratactic account clearly aspires to be. Suppose Galileo to have uttered 'It is raining' or 'I am hungry.' Would we regard

(18) Galileo said that it is raining.

and

(19) Galileo said that he is hungry.

as acceptable?

The point is even clearer with utterances of sentences with the verb in the continuous (sometimes—unhappily—called progressive) present (see note 6). If, during an earthquake, my friend utters 'The earth is moving' (that is, the earth beneath our feet, not the Earth), it would be bizarre to report him the next day by

(20) My friend said that the earth is moving.[12]

In these examples, tense-sequencing is forced, if we are to avoid reporting the speaker as having said something preposterous. If (18) is not acceptable, neither is its paratactic reformulation,

(21) Galileo said that. It is raining.

What makes the trouble I am alleging for the paratactic account is not the fact that the reference of pronouns or other referring expressions is determined by context, something others have worried about.[13] We can grant that an account of samesaying may be given that accommodates some kinds of indexicality and context-dependence. The problem is that the kind introduced by tense seems to make it impossible for a reporter to make the same utterance as was made by the reportee. To utter 'The Earth moves' today is not to say the same thing as what Galileo said in uttering that sentence in 1633, even if we interpret 'moves' as a habitual, our having good reason to do so notwithstanding. True—as noted above—an episodic reading is not available (in English) and true, we have good reason to believe that Galileo intended to assert what he took to be a law. Even so, we should not build into our report of what he *said*, as the paratactic account has us do, that it would be a law in 2019 that the Earth moves. Someone else uttering the same sentence may not have the same intention. Surely, though, he would have said the same thing as did Galileo.

Suppose Pliny to have uttered on the 21st of August, 79

(22) Vesuvius will erupt.

The tense-sequencing rule requires us to report him as in

(23) Pliny said that Vesuvius would erupt.

so as to avoid reporting him as saying something about what Vesuvius would be doing at times subsequent to our report. We need to do this to get the truth-

[12] Here 'moving' means something different than it does in Galileo's 'si muove'.

[13] Notably Blair (who also agrees with Davidson and Higginbotham in accepting (2) as the correct natural-language report) (2009: 33). On some views (e.g., Lepore and Cappelen) there is really no such thing as a context-free interpretation of a sentence.

conditions of his utterance right: he said something true as long as Vesuvius erupted at some the time between his utterance and our report, even if it never erupts again. But neither

(24) Pliny said that. Vesuvius will erupt.

nor

(25) Pliny said that. Vesuvius would erupt.

captures this, the first for the reason just seen, the second, because its content-sentence, being a conditional, cries out for completion ("if only...") and is, without that, ungrammatical.

It may be thought that the whole question of tenses can be finessed, and thus the propositions brought into line, by interpreting the verb in (1) as the habitual, as we do 'smoke' in 'Do you smoke?' That interpretation is, in fact, correct, but it is of no help in getting (1) to express the same proposition when uttered in 1630 and in 2019, respectively. We should not be understood as reporting Galileo as saying something about the Earth's habits in 2019, any more that we would want to report someone uttering 'Walter smokes' last year as saying something about Walter's habits today. This is so even if we are justified in believing that he would have been disposed to say then, and would say now, the same thing about the Earth's habits as he said in 1633.

The requirement that the reporter samesay the speaker is at the heart of Davidson's account. With the definition of samesaying in hand, he asks, what is needed if it is to be the case for Galileo's utterance and my report of it to satisfy it? His answer is that, unlike with quotational analyses of indirect discourse, which put the sentence uttered within the scope of 'said,' I need to actually say what it says; I need to use it, not merely mention it. Making the content clause an independent sentence accomplishes this: I can say it and say (with the first sentence) that *it* is what Galileo said. Here is what Davidson says: "If I merely *say* that we are samesayers, Galileo and I, I have yet to *make* us so; and how am I to do this? Obviously, by saying what he said; not by using his words (necessarily), but by using words the same in import here and now as his there and then" (1968:141).

In one sense of 'say,' of course, requiring the reporter to say (that is, assert) what his subject said would be absurd. Clearly, Davidson is not claiming that a reporter must himself assert what his subject did, that one cannot report without endorsing. Samesaying must be understood as limited to what Austin calls the locutionary act (1962: 94). It is a matter thus not of sameness of illocutionary act but only of sameness of sense and reference, with the latter being crucial. Sameness of illocutionary force is not required, hence my reporting what you asserted does not commit me to asserting what you did. This is evident in ordinary, tense-sequenced reports such as (6). The trouble for the paratactic account is that the problems of tense I have canvassed arise at the locutionary level, specifically with respect to reference. Tensed verbs ineliminably refer to different times—tense, we may say, determines reference. This is why utterances of sentences with a differently tensed verb express different propositions and their utterers say different things.

The underlying problem is that the paratactic account requires the content-sentence to do double duty. It has to be the vehicle both for the reporter's utterance and, albeit at one remove, for the utterance being reported on. No sentence can be both these things at once.[14] For this reason, one cannot really samesay the speaker one is reporting. Nor should one try. In saying what the speaker said, a reporter is not saying the same thing as the speaker did—to say what someone said is not to say it. In one way this is obvious, if saying is understood as a speech act, rather than just the uttering of a meaningful string. In uttering the second sentence in my report, I do not say what I say you said when you uttered it. Suppose I ask you to say what Galileo said. Am I asking you to say what it was that he said or to say it yourself? No doubt, the context will usually disambiguate. On the paratactic analysis, however, it is not clear that you can do the former except by doing the latter.

It is instructive to compare the case of saying, once again, with that of knowing. Setting tense aside for a moment, the same ambiguity is present with the latter. I can know what (=what it is that) you know without knowing what (=that which) you know, just as I can say what (it was that) Galileo said without saying what (=that which) he said (Austin 1946: 299).

I do not intend the analogy to be perfect. One difference is that I cannot say what you said without knowing what you said, as I can know what you know without knowing what you know.[15] What matters, though, is that if I know that which you know, we are, as we may put it, *sameknowers.* But with saying, the tensed verbs in the content-sentence of the original utterance and the content-sentence of the report, respectively, stand in the way of this.

The lesson is that the requirement of samesaying for correct indirect reporting is too strong. It asks that the original utterance and the utterance of the content-sentence of the report express the same proposition.[16] By contrast, reports that respect tense sequencing, such as (6), achieve the sameness of purport Davidson rightly seeks and which, ironically, the requirement of samesaying the paratactic analysis imposes frustrates. In a nutshell: I cannot samesay Galileo in the way proposed, no matter how I try to do it. If the paratactic account were right, I could not report what he said. But I can, too, report it, as in (6). And I can do so precisely because, contrary to the claim that is at the heart of the paratactic account, (6) cannot be parsed as two independent sentences. The relation between its main verb and its content clause, made explicit by sequence of tense, is all-important.

This also shows that accepting (2) as the correct natural-language report of (1) is a mistake. That it *is* accepted as such by almost everyone today may herald the imminent demise of the tense-sequencing rule in English. But as long as it

[14] For a discussion of similar problems with so-called mixed quotation, see Washington and Biro 2001.

[15] I could, if you spoke in a tongue unknown to me, by making a *direct* report. And I may do this even if you (appear to) speak in English but I am not sure—for whatever reason—that you are to be interpreted straightforwardly (Biro 1984, Washington and Biro 2001).

[16] Even if, as noted above, it allows for the sentences uttered to differ in meaning and perhaps even in truth condition (Higginbotham 1999: 207, Burge 1986:192).

has the rule, an account of indirect discourse in English should respect it, as the paratactic account does not.[17]

References

Austin, J.L. 1946, "Other Minds", in Urmson, J.O. and Warnock, G.J. (eds.), *Philosophical Papers*, 3rd ed., Oxford: Clarendon, 76-116.

Austin, J.L. 1962, *How to Do Things with Words,* Oxford: Oxford University Press.

Biro, J. 1984, "What's in a Belief?", *Logique et Analyse*, 107, 267-82.

Biro, J. 1992, "In Defence of Social Content", *Philosophical Studies*, 67, 3, 277-93.

Biro, J. 2011, "What is 'That'?", *Analysis*, 71, 4, 651-53.

Blair, D. 2009, "Bridging the Paratactic Gap", in Stainton, R.J. and Viger, C. (eds.), *Compositionality, Context and Semantic Values: Essays in Honour of Ernie Lepore*, Dordrecht: Springer, 31-58.

Burge, T. 1986, "On Davidson's 'Saying That'", in Lepore, E. (ed.), *Truth and Interpretation*, Oxford: Blackwell, 190-208.

Cappelen, H. and Lepore, E. 2005, *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*, Malden: Wiley-Blackwell.

Davidson, D. 1968, "On Saying That", *Synthese*, 21, 130-46.

Davidson, D. 1976, "Reply to Foster", in Evans, G. and McDowell, J. (eds.), *Truth and Meaning: Essays in Semantics*, Oxford: Oxford University Press, 17-80.

Davidson, D. 1999, "Reply to Elugardo", in Zeglen 1999, 114-15.

Elugardo, R. 1999, "Samesaying", in Zeglen 1999, 97-114.

Higginbotham, J. 1986, "Linguistic Theory and Davidson's Program in Semantics", in Lepore 1986, 29-48.

Higginbotham, J. 1999, "Tense, Indexicality and Consequence", in Butterfield, J. (ed.), *The Arguments of Time*, Oxford: Oxford University Press, 197-215.

Higginbotham, J. 2009, *Tense, Aspect, and Indexicality*, Oxford: Oxford University Press.

Larson, R.K. and Ludlow, P. 1993, "Interpreted Logical Forms", *Synthese* 95, 3, 305-55.

Lepore, E. (ed.) 1986, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.

Lepore, E. and Ludwig, K. 2003, "Outline for a Truth-Conditional Semantics for Tense", in Jokic, A. and Smith, Q. (eds.), *Tense, Time and Reference*, Cambridge, MA: MIT Press, 49-105.

McDowell, J. 1980, "Quotation and Saying That", in Platts, M. (ed.), *Reference, Truth, and Reality: Essays on the Philosophy of Language*, Boston: Routledge and Kegan Paul, 206-37.

[17] As noted above, not all languages have the kind of tense-sequencing rule English does. It must be decided case by case whether a given language has other resources for capturing the same distinctions or is to be judged as lacking in expressive power.

Segal, G., 1989, “A Preference for Sense and Reference”, *The Journal of Philosophy*, 86, 73-89.

Segal, G. and Speas, M. 1986, “On Saying ð∂†1”, *Mind and Language*, 1/2, 124-32.

Washington, C. and Biro, J. 2001, “A Logically Transparent Theory of Discourse Reporting”, *Mind and Language*, 16, 2, 146-72.

Zeglen, U.M. (ed.), 1999, *Donald Davidson*: *Truth, Meaning and Knowledge*, New York: Routledge.

Non-Doxastic Conspiracy Theories

Anna Ichino and Juha Räikkä***

** Università degli Studi di Milano – La Statale*

*** University of Turku*

Abstract

To a large extent, recent debates on conspiracy theories have been based on what we call the "doxastic assumption". According to that assumption, a person who supports a conspiracy theory *believes* that the theory is (likely to be) true, or at least equally plausible as the "official explanation". In this paper we argue that the doxastic assumption does not always hold. There are, indeed, "non-doxastic conspiracy theories": theories that have many supporters who do not really believe in their truth or likelihood. One implication of this view is that some debunking strategies that have been suggested to fight conspiracy theories are doomed to fail, since they are based on the false view that supporting a conspiracy theory means, *ipso facto,* believing in it—while they don't have grip in non-doxastic contexts.

Keywords: Conspiracy theories, Belief, Non-doxastic attitudes, Hope, Communication, Debunking strategies.

1. Introduction

In recent years, there has been a lot of discussion on why so many people support conspiracy theories, and on what, if anything, should be done to restrain the spread of conspiracist beliefs. Both the empirical debate on the possible causes of the popularity of conspiracy theories and the normative debate on how to deal with conspiracy theorists are usually based on what can be called the "doxastic assumption". According to that assumption, a person who supports a conspiracy theory *believes* that the theory is (likely to be) true or at least equally plausible as the "official explanation". This assumption is "doxastic", as it claims that *supporting* a conspiracy theory amounts to *believing* that the theory is (likely to be) true. According to the doxastic assumption, for instance, a person who supports the conspiracy theory that Princess Diana was murdered by the British Intelligence believes that this is so, or that it may very well be so. Her attitude towards the theory is doxastic in nature. (See e.g. Goertzel 1994; Sunstein and Vermeule 2009; Swami and Coles 2010; Wood, Douglas and Sutton 2012; Brotherton and French 2014; Coady 2012; Van Prooijen 2012; Van Prooijen and Acker 2015; Dentith 2016; Imhof and Lamberry 2017; Hagen 2018; Van Prooijen 2019.)

Argumenta (2021): 247-263
ISSN 2465-2334

DOI 10.14275/2465-2334/20200.ich

The doxastic assumption is natural and gets *prima facie* support from what people say and do. For instance, if a person openly defends the claim that the U.S. authorities must have known in advance that WTC towers would have been destroyed by referring to evidence concerning the normal practices of the U.S Intelligence, it seems reasonable to ascribe her a belief in the 9/11 conspiracy theory. However, we will argue that in some cases such belief ascription is questionable. That is, we will argue that there are what we call "non-doxastic conspiracy theories"—theories that have many supporters who do not really believe that their main claims are true or likely, as they have not considered the truth of those claims in the first place. The said theories are supported on non-doxastic bases.[1]

The view that supporters of conspiracy theories need not always believe in the theories they support is not completely new; it has been defended here and there (Ichino 2018; Hristov 2019). However, our discussion of the phenomenon is meant to be novel and revealing, in that it will examine in detail some of the psychological mechanisms underlying non-doxastic support for conspiracy theories, as well as the implications of the non-doxastic approach for practical interventions on those theories. In the first part of the paper, we will argue that (1) in some cases supporters of conspiracy theories merely *hope* that the theories they endorse are true, and that (2) in some other cases, by openly supporting those theories, they merely mean to *communicate* their support for the creators and the other supporters of those theories. We acknowledge that the evidence for these two models is not conclusive, and more empirical research is needed; but our argument shows that the two models deserve serious attention. On this basis, in the second part of the paper we will argue that those who are willing to debunk conspiracy theories should take the existence of non-doxastic conspiracy theories into account.

Importantly, our argument is not based on any especially controversial understanding of the nature of belief. We assume a minimal characterization of "belief" as a cognitive attitude involving the acceptance of some proposition as true—something that, at the functional level, amounts to displaying *at least some degrees* of sensitivity to evidence, holistic inferential integration with other doxastic states of the subject, and action-guidance. In so doing, we reject a "purely behavioral" (or "purely motivational") view according to which behavioral dispositions are not only necessary, but also sufficient for belief ascription. The characterization we adopt is widely accepted.[2]

[1] In epistemology, the notion of "doxastic theory" often refers to a theory according to which only beliefs can serve to justify beliefs; a "non-doxastic theory" is then simply a theory which denies that (Lyons 2009: 20). Our notion of "non-doxastic theory" is different. Our usage of the terms "doxastic" and "non-doxastic" is borrowed from debates in the philosophy of mind about the nature of phenomena like, for instance, delusions or confabulations (see e.g. Bortolotti and Miyazono 2015; Ichino 2018). In these debates, a doxastic theory is a theory according to which the phenomena in question involve a doxastic commitment (i.e. a *belief*) on the part of the subject—while non-doxastic theories deny that (arguing that subjects do not—or not always—believe the contents of their delusions and confabulations). In line with this usage of the terms, here we call "non-doxastic" those conspiracy theories that are not believed by the subjects who profess them.

[2] See e.g., among many others, Armstrong 1973, Velleman 2000, Williamson 2000, Bortolotti 2010, Ichino 2019. The specification 'at least to some degrees' is important,

Although the notion of belief is far from unproblematic in various other respects—which are at the center of lively debates in philosophy, psychology, anthropology, and elsewhere—we remain neutral with regard to many current controversies about it. For instance, we do not assume any particular stance on the debate on whether beliefs (as "on-or-off" attitudes) can be reduced to *credences* (that have degrees and correlate with the subjective probability that some proposition is true), or not (Jackson 2019; Levinstein 2019; Carter *et al.* 2019). Both belief and credence are doxastic attitudes, and in what follows we aim to show that such attitudes do not always play the role that they are commonly supposed to play in conspiracy theorizing. Similarly, we will not take any stance on the debate concerning *permissivism*—the view that the same body of evidence can justify more than one response, and that some beliefs can be merely permissible (rather than obligatory) in the face of evidence (Ballantyne 2018; Schultheis 2018; Axtell 2019). Permissivism is a normative doctrine, and our point here is mostly descriptive: we aim to describe the mechanisms underlying support for conspiracy theories.[3]

To begin, let us start by defining the notion of conspiracy theory.

2. What Are Conspiracy Theories?

The definition that follows is meant to clarify the discussion; we do not mean to suggest that it is the only appropriate way to use the concept. By "conspiracy theory" we indicate an explanation of a given event that: (1) refers to actual or alleged conspiracies or plots (*Conspiracy Criterion*); (2) conflicts with the received explanation of the said event, providing an alternative to the "official view" of that event (*Conflict Criterion*); and (3) offers insufficient evidence in support of the alternative explanation, so that it is not considered as a competitive scientific theory or anything like that (*Evidence Criterion*). These criteria are meant to be necessary and jointly sufficient for something to count as a conspiracy theory.

So, for instance, the claim that there was a Cuban plot behind the murder of President John F. Kennedy is a conspiracy theory as it explains a political event by referring to a conspiracy and offers an alternative to the official view. The theory blames one group for conspiring (Cubans) and another group for failing to notice it (the U.S. authorities). In many cases, the group that is accused of hatching a conspiracy consists at least partly of the people who should know and tell the truth. For instance, the claim that genetically modified food kills people and the authorities know it (but do not tell it), is a conspiracy theory that blames authorities both for scheming and concealment. The theory is supposed to explain why some business secrets are kept as such.

since we all know that in limited cognitive agents like us, belief's sensitivity to evidence, inferential integration, and action-guidance might not be perfect. Note also that, on the characterization we are proposing, believing something does not imply that the person is aware of her belief; conversely, a person's conviction that she has a certain belief does not imply that she actually has it.

[3] Our discussion concerns *sincere* supporters of conspiracy theories. Some people may disseminate conspiracy theories just because they benefit from the large acceptance of such theories, although they are aware that they do not believe in such theories at all, and do not sincerely support them in any way. These are not the sort of people we are concerned with.

Let us look at the *Conspiracy Criterion*, the *Conflict Criterion* and the *Evidence Criterion* more closely.

The *Conspiracy Criterion* is based on the idea that reference to a conspiracy is a necessary condition for an explanation to count as a conspiracy theory. If an alternative explanation of a given event does not refer to a plot or a conspiracy, then it is not a conspiracy theory, however denialist the explanation may otherwise be. This raises the question of what counts as a "conspiracy". For the sake of this discussion, "conspiracy" can be defined as a concealed collective activity whose aim or nature conflicts with the so-called positive morality (which reflects our present moral commitments) or with *prima facie* duties, especially if the goal of the activity differs from the goals that its promoters are authorized to pursue.[4] Secret plans to organize birthday parties are not conspiracies, as their aim does not conflict with morality. Secret military operations are not usually called conspiracies, as far as they have an authorized goal. An example of a conspiracy is the Volkswagen Group's decision to lie about the emissions of their cars and deceive their customers. It was a carefully designed secret plan (that was collectively realized) which clearly conflicted with *prima facie* duties, including a duty not to (plan to) deceive people. The Group was not authorized to cheat on the consumers.

The *Conflict Criterion* is meant to separate conspiracy theories from other sorts of theories that refer to conspiracies. There are hundreds of historical accounts that mention "conspiracy" as a part of the explanation of a historical event, but they do not count as conspiracy theories on our view, as far as they represent the received view of history (Keeley 1999: 116; Levy 2007: 187; Räikkä 2018: 211). The claim that Bolivian authorities conspired with the CIA to kill Ernesto Che Guevara in 1967 is not a conspiracy theory, but the "official" truth about Che Guevara's death. An explanation that refers to a conspiracy is a conspiracy theory only if the relevant epistemic authorities, more or less unanimously, find the conspiracy claim strikingly implausible, or would find it strikingly implausible in case they considered it. The view that vaccines will kill millions of people and health authorities know it (but do not confess it) is a conspiracy theory, as it is strikingly implausible according to the epistemic authorities on which we normally rely—such as the scientific community, mainstream media, investigative journalists, various state authorities and agencies, and so on.

The *Evidence Criterion* helps to distinguish between conspiracy theories and some historical theories that may also refer to conspiracies and conflict with the received view. For instance, the claim that Rasputin was killed by the British intelligence service is not considered (or is not always considered) to be a conspiracy theory, but a competitive historical theory about the death of Rasputin. Those two kinds of theories—conspiracy theories and (what we can call) "minoritarian" theories that refer to conspiracies—differ with respect to the quality of the evidence they provide. Conspiracy theories offer relatively little (good quality) evidence in support of the conspiracies they talk about; while minoritarian scientific or historical theories, which may likewise make claims about con-

[4] The second disjunct is needed because it is easy to imagine cases in which conspiring is morally acceptable, all things considered. There are many historical examples of morally acceptable conspiracies. Operation Valkyrie (the secret plan to kill Hitler) is an obvious example here.

spiracies, offer a good amount of good quality evidence in support of their claims. They may not convince most of the experts, but they are taken seriously, because of the evidence they provide. The quality of the sources that are used in conspiracy theories is not as good (cf. Harris 2018: 243; Levy 2019: 70).[5]

This definition of the notion of conspiracy theory has several merits. First, the definition reflects relatively well the ordinary usage of the term and seems to be extensionally adequate. When people talk about "conspiracy theories", they usually refer to claims that blame a given group of people for conspiring and that go strongly against the received view. And the examples we can think of theories that are commonly classified as conspiracy theories would count as such according to our definition. Second, our definition does not imply that conspiracy theories must be false. Epistemic authorities make mistakes—although it is important to notice that usually we know about such mistakes because epistemic authorities themselves have produced the information that helps us to notice them. Third, the definition does not imply that those who represent epistemic authorities could not be conspiracy theorists. A biologist, a journalist or a historian, may well present an explanation which refers to an alleged conspiracy, but which is pure non-sense according to most others.[6] Fifth, by virtue of the *Conflict Criterion*, our definition makes the notion of conspiracy theory a relative one (i.e., relative to different historical contexts) given that epistemic authorities may change their views over time, and so also what conflicts with such views will change accordingly. This is an advantage, because something that counts as a conspiracy theory today may turn out not to be such anymore, in the light of new evidence that might emerge.[7] Finally, our definition does *not* imply that new theories that are not (or have not yet been) evaluated by the relevant epistemic authorities cannot be genuine "conspiracy theories". On our view, a new theory that refers to a conspiracy is a conspiracy theory, if the epistemic authorities *would* find the conspiracy claim strikingly implausible, *after considering it*.

3. Non-Doxastic Support for Conspiracy Theories

As we acknowledged, there are *prima facie* reasons to take people's attitude towards conspiracy theories to be doxastic: after all, people often give sincere verbal assent to such theories—and we generally take sincere verbal expressions of assent as a guide to belief ascription. On closer inspection, however, there are also reasons that speak *against* belief ascription here. Alleged beliefs in conspiracy theories are commonly taken to be irrational to relevant degrees, due to their weird contents that conflict with the views of widely recognized epistemic authorities. Surely, we should avoid ascribing irrational or epistemically irrespon-

[5] Importantly, the claim that conspiracy theories are weakly supported by evidence does not imply that they are false: poor evidential support is compatible with truth—and some conspiracy theories do indeed turn out to be true.

[6] Obviously, epistemic authorities do not form a monolithic body, and may well disagree with each other on various issues.

[7] The view according to which there was a Jewish conspiracy against Christians was an official truth in Germany during World War II, and those who endorsed that view were not (always) taken to support a conspiracy theory. However, now we can say that many Germans at that time supported a conspiracy theory concerning the Jews. The reason why we can say this is that today the claim conflicts with the received view of history. This is why we now count it as a conspiracy theory (Räikkä 2018: 211).

sible beliefs to each other, if there are alternative mental states ascriptions available that make sense of each other's behavior without involving irrationality or irresponsibility (or involving less of them). This suggests that we should take the non-doxastic option seriously, and consider possible mechanisms that may lead one to express support for a conspiracy theory while actually *not believing* that the theory itself is correct or likely.

Here we will identify two such mechanisms—mechanisms of non-doxastic endorsement—and we will consider empirical studies that support the idea that such mechanisms are indeed at play in a number of cases of conspiracy theories advocacy. We will call the first mechanism the "Hope Process" and the second mechanism the "Communication Process". We will introduce both of them by describing imaginary examples that are not directly related to conspiracy theories. Then we will argue that something similar to what happens in such examples may happen when a person endorses a conspiracy theory without really believing it. Notice that our point here is programmatic: we sketch two models that, if proven true, would have important implications. But we also provide some initial evidence for their truth, thereby indicating avenues for future research.

3.1. The Hope Process

Consider a high-school drama. There is a lucky guy in the school who gets relatively good grades, is good at sports, and gets attention from his colleagues. There is another guy in the school who is not as lucky as the lucky guy, and who envies the lucky guy, although he does not realize it, because of his poor self-knowledge. One day the lucky guy does not say "hello" to the unlucky guy, even if they know each other well enough. Not saying "hello" is an accident, but the envious unlucky guy has a different explanation. At first, he is just angry; but after a couple of days he is sure that the lucky guy is a selfish, arrogant, and untrustworthy person. That is why the lucky guy does not even say "hello" to him. However, the unlucky guy deceives himself. The belief that there is something seriously wrong with the lucky guy helps him psychologically. The evidence in favor of such belief is weak and inconclusive—what really supports it are motivational mechanisms of self-enhancement: now he can think that, actually, he, and not the other boy, is the clever guy. The unlucky guy starts to disseminate strange claims—whenever it is possible and fits the social situation. For instance, he claims that the lucky guy typically does not keep his promises, and that the lucky guy often lies. Given the unlucky guy's view of the lucky guy, these claims make sense. Untrustworthy people do not always keep their promises and they can lie every now and then.

Then one day someone from the school tells the unlucky guy that actually the lucky guy usually keeps his promises. She has a plenty of evidence for that. The unlucky guy does not really question the said evidence, but he simply replies that, in any case, the lucky guy is a liar. For him, it is not important to insist that any *particular* dismissive claim about the lucky guy is true. It is enough that some, or at least one, of them is true. He believes that the lucky guy is selfish, arrogant, and untrustworthy; and it is psychologically important for him that this belief is correct. This belief predicts that at least one dismissive claim is true, and therefore he really *hopes* that they are not all false. Whenever he considers one of them, he hopes that it is true. But he does not truly *believe* any of them, although he does not think that they are false either. When he says what he says, he is not lying.[8]

Now, in some cases something similar may be going on when a person supports a conspiracy theory. It need not be the case that the person really believes in the theory. It may be that she merely hopes that the theory holds, as the theory supports some more general view which she is motivated to believe—such as, for instance, the general view that the "authorities" or "establishment" (i.e. the State, the scientific community, the business companies, the media, and so on) are, in general, seriously unreliable and untrustworthy. Her motivations to believe this general view may be rooted in her poor social conditions and overall unsatisfaction with her life. She may be unemployed, down and out, badly disappointed by the "system", and lacking sense of control over her life (cf. Abalakina-Paap *et al.* 1999; Uscisnki and Parant 2014). Believing that the "system" itself is untrustworthy might well provide some comfort to her: meaning that her problems are largely caused by others rather than by herself. Due to her motivation to hold a belief in this general view, she may hope that the conspiracy theories that support this view are true.[9] When a conspiracy theory she supports is shown to be rubbish, she does not care about the issue too much, but simply shifts to another conspiracy theory, since some (or at least one) of them *must be true*. This is psychologically important. When she disseminates those conspiracy theories, she is not lying, as she does not consider them to be false. She has simply not considered them from an epistemic point of view. What she has considered, albeit in a motivationally biased way, is the general view that all main institutions are untrustworthy. When she disseminates the conspiracy theories predicted by this general view, she may think that she is doing something important.

Hoping and believing are different—and typically incompatible—things. If a person consciously and openly believes that something is the case, arguably she cannot hope for that thing (anymore), since hope is accompanied with uncertainty. Hoping and wishing, too, are different things. A person can wish that she could jump into the moon even if she thinks that it is impossible. But she cannot hope it, if she thinks that it is impossible. Thus, a person who hopes that a conspiracy theory is true does not believe that it is impossible that it is true.[10]

[8] This might be an instance of what Harry Frankfurt (2005: 55-56) calls "bullshit". As he writes: "It is impossible for someone to lie unless he thinks he knows the truth. Producing bullshit requires no such conviction. A person who lies is thereby responding to the truth, and he is to that extent respectful of it. When an honest man speaks, he says only what he believes to be true; and for the liar, it is correspondingly indispensable that he considers his statements to be false. For the bullshitter, however, all these bets are off: he is neither on the side of the true nor on the side of the false". Notice, however, that a person who expresses her support for a conspiracy theory says something that has (for her) a clear *function*. Her sentences are not irrelevant, although their truthfulness is not crucial for her.

[9] The argument here is not that people have a motivation to believe in conspiracy theories. Our claim is that there is a basic motivation to think that main social institutions are not reliable. This "thinking", in turn, may or may not be doxastic: it can take the form of a *belief*—like the belief that "the establishment is untrustworthy", but also the (purely) affective form of a *distrust* towards the "establishment".

[10] See Bovens 1999; Meirav 2009; Miceli and Castelfranchi 2010; Govier 2011; Martin 2011; Kadlac 2015.

Importantly, a person who does not believe in a conspiracy theory but merely hopes that it is true would not typically say that she is only hoping, if asked. She would rather say that she really believes in the theory. What is at stake here, on our view, is a special sort of meta-cognitive mistake. She does not believe, but rather merely hopes, that the theory is true; but she mistakenly takes her hope to be a belief.[11] There may be various reasons why she makes this mistake. For one thing, hoping that a conspiracy theory is true would be an instance of hoping something *bad*, and hoping something arguably involves wanting that thing. But most of us think that wanting bad things to be true is not appropriate—so, hoping them would not be appropriate, either. Moreover, psychologically speaking, it seems important for her to believe that she *believes* the conspiracy theory, rather than merely *hoping* it—given that belief is the appropriate attitude towards things that are true, and she does indeed hope the conspiracy theory to be true. Hence the mistaken belief self-ascription.

This model of the mechanisms underlying the commitment to conspiracy theories is non-doxastic in that it denies that such commitment amounts to believing the theories in question. On this model, supporting a conspiracy theory amounts to hoping, rather than believing, that the theory is true. On the other hand, the model credits conspiracy theories' supporters with some other more general *beliefs*—namely, beliefs about the untrustworthiness of the "system"—which, in turn, explain their hopeful commitment to the conspiracy theories themselves. Supporting a specific conspiracy theory may then be seen as an indirect way to express a deeper more general conviction.[12]

Empirical studies on conspiracy theories suggest that something like the Hope Process just described is not unlikely. There is a good amount of empirical evidence that people who support conspiracy theories do not trust the "authorities" and the "establishment" as much as those who are not eager to endorse conspiracy theories (Swami and Coles 2010; Swami 2012). There is also some evidence that people who support conspiracy theories sometimes have "personal reasons" (that is, a motivation) to adopt the general claim that the main social institutions are untrustworthy and unreliable (Goertzel 1994; Douglas and Sutton 2011).[13] Furthermore, there is empirical evidence that if a person supports one conspiracy theory, this increases the probability that she will adopt another conspiracy theory as well (Swami and Coles 2010; Lewandowsky 2013). This result is well in line with the dynamics of the Hope Process.

Finally, there is empirical evidence that people are willing to support conspiracy theories whose claims conflict with each other and that cannot all be true at the same time (Wood, Douglas and Sutton 2012). These findings are due to Karen Douglas and her group, who interpret them within a doxastic frame-

[11] It is not uncommon that a person ascribes herself beliefs that she does not actually have, or that she does not ascribe herself beliefs that she actually has (Räikkä and Smilansky 2012).

[12] In fact, as we noted, we are open to the possibility that also such a general conviction might take non-doxastic forms – involving an affective attitude of distrust, rather than a full-fledged doxastic attitude of belief (see footnote 9 above). Our point here is that, *even granting* that a subject's general conviction about the untrustworthiness of the 'establishment' is a belief, her attitude towards the specific conspiracy theories that she endorses as a result of that general conviction might well be a non-doxastic attitude, instead.

[13] According to Goertzel (1994: 731), "belief in conspiracies was correlated with anomia, lack of interpersonal trust, and insecurity about employment".

work according to which conspiracy theories supporters hold openly contradictory beliefs, thereby incurring in blatant irrationality. As some authors have pointed out, however, this interpretation is strikingly uncharitable.[14] The Hope Process provides a much more charitable interpretation: a person can certainly adopt conflicting conspiracy theories when she does not actually *believe* in them, but merely *hopes* that some, or at least one, of them is true. Having such hopes does not involve any contradiction, and there are no psychological mysteries here—nor indeed irrationality.[15] While granting that the empirical data just mentioned *might* be explained also within a doxastic framework, we observe that the non-doxastic framework provided by the Hope Model has some clear advantages here.

3.2. The Communication Process

Suppose that a young person would like to identify herself as a part of the growing popular movement that opposes the use of plastic products. She is deeply concerned about environmental issues and would like to flag her attitude by supporting the anti-plastic movement. The leaders of this movement disseminate their message in their websites and in social media. The person who would like to be involved forwards these messages, although often she does not really understand their content. After all, they include rather complex claims about chemistry and biology—claims that are not common knowledge.[16] Sometimes it happens that a claim of the movement is publicly shown to be false (by the relevant experts). But that does not really perturb the person who continues to disseminate the movement's newsletters. The key issue for her is expressing agreement rather than establishing truth. She would like to show that she supports the movement, and disseminating the messages is her way to *communicate* that. By disseminating the claims of the movement, she does not aim to say that the claims are true. She merely wants to express her participation and commitment to the movement's general agenda, which she takes to be important and admirable. Her support for the messages is basically an indirect way to show this more general commitment.

Now, it may be that in some cases something similar happens when a person expresses her support for a conspiracy theory. She needs not believe the theory at all; simply, since she admires the people who support that theory, she

[14] According to Basham (2017: 64): "Wood et al.'s interpretive mistake is so surprising because it is so clear. Simply, the researchers conflate participants' reports of strong suspicions with settled beliefs".

[15] One here might wonder whether hoping mutually inconsistent propositions isn't actually irrational, just like believing mutually inconsistent propositions is. But this doesn't seem to be the case. Although the question of what precisely makes one's hope that *p* rational is complex and debated, indeed, it seems clear that hope undergoes different (and arguably looser) rationality constraints than belief. According to Meirav (2009), for instance, the rationality of one's hope about a given outcome depends on the rationality of her belief about the "goodness" of an external factor upon which the realization of that outcome causally depends. On a view like this, given that the same external factor may be responsible for the realization of mutually inconsistent outcomes, hoping for mutually inconsistent outcomes would not be *ipso facto* irrational. We are grateful to an anonymous referee for raising this issue.

[16] An interesting question here is in what sense one can "believe" propositions that she does not (or not fully) understand (see Recanati 1997).

might want to express her support for them by disseminating their claims. A person who is concerned about the risks of vaccination may very well support a conspiracy theory developed by a group who thinks that vaccination is riskier than it is commonly taken to be, and much riskier than the relevant epistemic authorities publicly admit. By expressing support for that particular conspiracy theory, a person needs not really believe it, as her point is merely to flag the opinion that the group has an important agenda and that she therefore stands by them.

When she disseminates the conspiracy theory on social media, she thinks that she is doing something important—namely, pointing out that the issue of vaccination safety is worth attention. But her relation to the content of that theory does not involve a doxastic commitment. She supports it merely because of pragmatic reasons. In so doing, she does not lie, for she does not think that the theory is false. Its truthfulness is not an issue that concerns her. If the theory turns out to be false, this would not be the end of the world. The relevant group may have another conspiracy theory or some other radical claim to which she can shift to communicate her agreement with them. Here again, as in the Hope Process, a person who supports a conspiracy in this way might not be aware that she does not really *believe* in the theory; she might simply lack a clear view about what her attitude towards the theory she disseminates is.

Empirical research on conspiracy theories suggests that conspiracy theorizing and support for conspiracy theories are often politically motivated (Fenster 1991; Knight 2002; Uscinski and Parent 2014; Cassam 2019). Both psychological and historical studies show that a person's political views are clearly connected to conspiracy theorizing, in particular, to the contents of the relevant theories (Olmsted 2009; Douglas and Sutton 2015).[17] Jaron Harambam (2017: 185) has observed that the "activism of the conspiracy milieu can be understood as a form of 'subpolitics'—a bottom-up form of politics outside of the formal political arena". These results are well in line with the dynamics of the Communication Process. When the aim of the person who disseminates and defends a conspiracy theory is merely to communicate her more general political identity, she needs not believe in the specific details of the theory (although of course she *might* believe in them). If a person supports a conspiracy theory in this way—i.e. merely as a mean to express her broader political views—again, the doxastic assumption does not hold.

Suppose that someone disseminates a no-vax conspiracy theory only in order to communicate that in her view vaccination safety needs more attention, and those who seek to defend the "right to choose" are good people. The person who disseminates the theory is part of the social process in which false claims spread.[18] Of course, it is possible that the person's audience understands that she

[17] Douglas and Sutton (2015: 101) argue that a feature of "climate change conspiracy theories is that they appear to be politically loaded, dividing opinion according to people's position on the spectrum between right and left. With the right wing emphasizing the production of wealth rather than its redistribution, and opposing governmental regulation and interference, it is not surprising that right-wing political identification is associated with disbelief in climate change".

[18] We say "false" here given that conspiracy theories conflict with the views that the relevant epistemic authorities more or less unanimously accept, so generally there is good reason to take them to be false. Still, as we noted, they might *at least in principle* be true—since bad justification is compatible with truth (see footnote 5 above).

cannot really mean what she says (about the alleged conspiracy), and she is just trying to make the point that some issues concerning vaccination should be more seriously discussed. In a case like this, the audience would know that the person does not truly believe the conspiracy theory, but expresses her support for it merely for communicative reasons; hence less harm would result. But presumably this is not, as a matter of fact, what typically happens most of the time.

Importantly, the two processes just sketched—Hope and Communication—must not be alternative to each other, but may also work in conjunction. *Hoping* that a certain conspiracy theory is true and seeking to *communicate* your support for the advocates of such theory may well go hand in hand. And, indeed, we can observe the same basic structure in both processes: the apparent belief in a given conspiracy theory actually amounts to endorsing something else.

Our argument for the psychological reality of those non-doxastic processes so far has been mainly abductive: we have argued that assuming those processes to be at play can explain a range of empirical data—and it can do that more charitably than some popular alternative explanations do. We now turn to some implications of our non-doxastic approach—implications which, as we shall see, provide a critical testing ground for the approach itself.

4. Implications for Debunking Strategies

We have argued that there are *non-doxastic* conspiracy theories—conspiracy theories that are not really *believed* by all of those who support them. The fact that someone expresses support for a conspiracy theory is not a sufficient reason to attribute to her a *belief* that the theory is true or likely. We have argued that in some cases supporters of the conspiracy theories merely *hope* that the theories they endorse are true (the Hope Process); and that in some cases they simply mean to *communicate* their support for the other supporters and disseminators of those theories (the Communication Process). Our claims get support from various empirical and historical studies, the results of which are nicely understood in the light of non-doxastic theory acceptance.

We will now consider some implications of our argument for possible debunking strategies. Those who are willing to debunk conspiracy theories, we will argue, should take the possibility of non-doxastic conspiracy theories into account when designing their practical interventions. Our point here is not to argue that debunking is a good idea.[19] We only argue that if someone finds the idea attractive, then she should understand what she opposes. If belief is not the attitude that is involved in supporting conspiracy theories, the game changes. Let us consider two examples of debunking suggestions. They are both problematic, if applied to non-doxastic conspiracy theories.

4.1. First Debunking Strategy: Adding Cognitive Diversity

It is often argued that one of the factors that make some people believe in conspiracy theories is their imperfect epistemic environment. Most people live in “epistemic bubbles” and, unfortunately, some bubbles tend to be conspiracy theory friendly to a considerable degree. In order to fight against the spreading

[19] The view that conspiracy theories require counter action is rather common. For a public defense of such view, see e.g. Bronner *et al.* (2016: 29).

of conspiracy theories, on this view, people should therefore try to increase the *cognitive diversity* of the groups who suffer from one-sided information that favors conspiracy theories.

This idea can take extreme forms. So, for instance, Cass R. Sunstein and Adrian Vermeule (2009: 219-220) famously argued that "cognitive infiltration of extremist groups" would "undermine the crippled epistemology of believers by planting doubts about the theories and stylized facts that circulate within such groups, thereby introducing beneficial cognitive diversity". The "limited informational environment" of conspiracy theorizers should be made more open and diverse—if necessary, by means of secret governmental operations (Sunstein and Vermeule 2009: 210, 218).[20] In his book on *Conspiracy Theories and Other Dangerous Ideas* Sunstein (2014: 32) stresses the point once again: if necessary, the state should conspire against citizens. The idea of fighting against conspiracy theories by adding cognitive diversity needs not take these extreme forms, though. Surely one can try to improve people's epistemic environments by various means, including means that are consistent with democratic values (and more likely to achieve their end).

But the strategy of increasing cognitive diversity is based on a doxastic assumption. And, as we have argued, this assumption is not always correct: there are likely to be non-doxastic conspiracy theories that are not *believed* by their supporters. Increasing cognitive diversity is unlikely to influence a person who supports a conspiracy theory only in the sense that she hopes that the theory is true (the Hope Process). Even if her epistemic environment were more or less perfect in terms of having a diversity of points of views, she could still *hope* that the conspiracy theory she supports is true. On our model, the relevant hope is grounded in a more general motivated belief—and increasing cognitive diversity is not likely to shake that general belief. Similarly, increasing cognitive diversity is unlikely to influence a person who endorses a conspiracy theory just in order to express her support for some group or movement (the Communication Process). Expressing support is a pragmatic reason that will not be displaced by increased cognitive diversity. Thus, if a person would like to debunk conspiracy theories and considers the policy of increasing cognitive diversity as a mean, she should first make sure that she is not dealing with a non-doxastic conspiracy theory. For if she is, the strategy might not be very effective.

4.2. Second Debunking Strategy: Teaching Logical Thinking

It has been argued that people who support conspiracy theories have defective logical competences and fall pray of various formal and non-formal fallacies. Robert Brotherton and Christopher C. French (2014: 246), for instance, argued that "conspiracy theories, similarly to other anomalous beliefs, are associated with reasoning biases and heuristics". An example here is the conjunction fallacy, to which people who endorse conspiracy theories seem to be "particularly susceptible" (Brotherton and French 2014: 246). A person who commits the conjunction fallacy thinks that the probability of two events occurring together is larger than the probability of either of them occurring alone—which, of course, cannot be true. A person who believes that there is 20% likelihood that "It rains tomorrow" should not believe that there is 30% likelihood that "It rains and

[20] For a criticism, see e.g. Hagen 2010; Hagen 2011; Coady 2018.

winds tomorrow". If conspiracy theorizing arises from bad reasoning, then those who would like to fight against the spread of conspiracy theories should try to improve people's logical skills, or their critical and scientific thinking more generally.

This suggestion, however, may have limited validity. Although the policy of educating people sounds good in general and would most probably have some desirable effects, this strategy is not likely to work in the context of non-doxastic conspiracy theories. As per the Hope Process, a person who hopes that a particular conspiracy theory is true (as its truth would strengthen her overall worldview) may not be that interested in the logical grounds of the theory. Indeed, hoping does not undergo the same normative constraints as believing. While the propositions we believe ought (at least ideally) to be integrated with each other into a logically consistent whole, there is nothing wrong in hoping that a given proposition is true even if it is not logically connected to other propositions that we take to be true. Hoping is possible until its object is considered demonstrably impossible.

Similarly, in the Communication Process, a person who uses a conspiracy theory merely as a means to communicate her ideological stance needs not be too much concerned about the logical grounds of the theory she refers to. So, improving her logical skills will not help much in fighting her penchant for conspiracist thinking. Again, all this suggests that if we would like to debunk conspiracy theories, we should first check whether we are dealing with theories that are supported non-doxastically. A person can support a conspiracy theory non-doxastically even if her logical skills are more or less perfect.

Of course, although many debunking strategies are based on a doxastic assumption, the view that the dissemination of conspiracy theories should be opposed must not, in itself, be based on that assumption. Indeed, one might argue that even if people's attitudes towards conspiracy theories are non-doxastic, insofar as those attitudes influence people's actions and reactions, leading to potentially dangerous behavior, they should be somehow "debunked". Although most philosophers think that people are free to speculate about possible conspiracies and to disseminate such speculations, the issue of whether and how the said speculations should be restrained becomes more and more pressing. In relation to the approach we defended here, then, the question arises of what should be done if a clearly harmful and mistaken conspiracy theory (say, an anti-Semitic theory) is supported mainly on non-doxastic grounds. What we have argued suggests that a promising way to oppose such theories might pass through policies aimed that enhancing people's trust in major social institutions—perhaps with the help of political programs that make the institutions more transparent and accountable. While a proper development of this suggestion goes beyond the scope of our present discussion, however, our aim here was more general: we meant to show that *whatever one might want to do* of conspiracy theories, she should first get clear on the mechanisms that underlie them.

Importantly, as we noted, the implications of our non-doxastic account for different debunking strategies may also provide a critical testing ground for the account itself. Insofar as the account predicts the failure of a given debunking strategy, indeed, once that strategy is put into place it will be possible to check whether or not the prediction is confirmed. Although successes and failures in this area are not always easy to assess, then, the non-doxastic model that we defended in this paper is, at least in principle, susceptible of empirical confirmation.

5. Concluding Remarks

To a large extent, the academic discussion on conspiracy theories has been based on the doxastic assumption. According to that assumption, a person who supports a conspiracy theory has a *belief* concerning it. We have argued that the doxastic assumption does not always hold, and that the results of empirical studies support the suggestion that there are "non-doxastic conspiracy theories"—theories that are not really believed by their supporters. We introduced two ways in which a person may support a conspiracy theory without really having the relevant beliefs about it. First, she may hope that the theory is true, as its truth would strengthen a more general worldview that is psychologically important for her. Second, she may express her support for the theory in order to express her political and ideological commitments, even if she has not really considered whether the theory is true. Many debunking strategies assume that people who support conspiracy theories have *beliefs* about them, and such beliefs should therefore be the targets of the relevant debunking interventions. But if what is at stake are not actually false beliefs and defective epistemic environments, then the relevant interventions should be redirected.

It is worth emphasizing again some implications of the view we defended for the assessment of the rationality of people's attitudes towards conspiracy theories. The charge of irrationality that is generally raised against such attitudes is based on the doxastic assumption—the point being that it is irrational to believe in conspiracy theories which are badly supported by the relevant epistemic authorities. But insofar as the doxastic assumption is questioned, the charge of irrationality may be reconsidered as well. As we noted, hope is not governed by the same epistemic norms that govern belief. And one may have good reasons to hope that a given conspiracy theory is true. Similarly, there is nothing especially irrational in communicating one's position by saying something different from what one wants to communicate: that sort of use of language is indeed common, although it may and does lead to confusions.[21]

This said, it is also worth noting that conspiracy theories supporters are likely to display some sort of irrationality at least at a meta-cognitive level—due to their unawareness about the non-doxastic status of their own attitudes. Indeed, we have seen that those who support conspiracy theories non-doxastically are often unaware that they do not really believe those theories: their self-knowledge is somewhat faulty, in the motivationally biased way we described—which is a far from ideal epistemic situation.

Last but not least, note that saying that attitudes towards conspiracy theories might be less *epistemically* irrational than they are often taken to be does not amount to saying that there is *nothing whatsoever* wrong with them. At the very least, such attitudes can be *morally* problematic, insofar as they involve accusations which are not well-supported by evidence. People who disseminate conspiracy theories without really believing them seem disturbingly unconcerned about truth and somewhat immune to normal evidential criteria. Surely, one

[21] The claim that people's attitudes towards conspiracy theories might not be irrational after all—or, anyway, that they might be less irrational than we commonly think—has recently been defended also by Levy 2019, within a doxastic framework where the relevant attitudes are taken to be beliefs.

should worry about truth and evidence if she is going to spread blame and accusations against other people.[22]

References

Abalakina-Paap, M. *et al.* 1999, "Beliefs in Conspiracies", *Political Psychology*, 20, 637-47.

Armstrong, D.M. 1973, *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.

Axtell, G. 2019, "Well-Founded Belief and the Contingencies of Epistemic Location", in Bondy, P. and Carter, J.A. (eds.), *Well-Founded Belief*, London: Routledge.

Ballantyne, N. 2018, "Is Epistemic Permissivism Intuitive?", *American Philosophical Quarterly*, 55, 365-78.

Basham, L. 2017, "Pathologizing Open Societies: A Reply to the *Le Monde* Social Scientists", *Social Epistemology Review & Reply Collective*, 6, 59-68.

Basham, L. 2018, "Joining the Conspiracy", *Argumenta*, 3, 2, 271-90.

Bortolotti, L. 2010, *Delusions and Other Irrational Beliefs*, Oxford: Oxford University Press.

Bortolotti, L. and Kengo, M. 2015, "Recent Debates on the Nature and Development of Delusions", *Philosophy Compass,* 10/9, 636-45.

Bovens, L. 1999, "The Value of Hope", *Philosophy and Phenomenological Research*, 59, 667-81.

Bronner, G. *et al.* 2016, "Let's Fight Conspiracy Theories Effectively", *Le Monde*, June 6th.

Brotherton, R. and French, C.C. 2014, "Belief in Conspiracy Theories and Susceptibility to the Conjunction Fallacy", *Applied Cognitive Psychology*, 28, 238-48.

Cassam, C. 2019, "Why Conspiracy Theories Are Deeply Dangerous", *New Statesman*, October 7 (online).

Carter, J.A., Jarvis, B. and Rubin, K. 2019, "Belief Without Credence", *Synthese* (online).

Coady, D. 2012, *What to Believe Now: Applying Epistemology to Contemporary Issues*, Singapore: Wiley-Blackwell.

Coady, D. 2018, "Cass Sunstein and Adrian Vermeule on Conspiracy Theories", *Argumenta*, 3, 2, 291-302.

Connor, C. and Weatherall, J. 2019, *The Misinformation Age: How False Beliefs Spread*, London: Yale University Press.

Dentith, M. 2016, "When Inferring to a Conspiracy Might Be the Best Explanation", *Social Epistemology*, 30, 572-91.

[22] A version of this paper was presented at a University of Milan *Seminar of Mind and Language* in February 2020; we are grateful to the audience of that event for their helpful feedback. We also wish to thank two anonymous referees of this journal for their really insightful comments, as well the journal editors for their generous work.

Douglas, K.M. and Sutton, R.M. 2011. "Does It Take One to Know? Endorsement of Conspiracy Theories Is Influenced by Personal Willingness to Conspire", *British Journal of Social Psychology*, 50, 544-52.

Douglas, K.M. and Sutton, R.M. 2015, "Climate Change: Why the Conspiracy Theories Are Dangerous", *Bulletin of the Atomic Scientists,* 71, 98-106.

Fenster, M. 1991, *Conspiracy Theories: Secrecy and Power in American Culture*, Minneapolis: University of Minnesota Press.

Frankfurt, H.G. 2005, *On Bullshit*, Princeton: Princeton University Press.

Goertzel, T. 1994, "Belief in Conspiracy Theories." *Political Psychology* 15, 731-42.

Govier, T. 2011, "Hope and Its Opposites", *Journal of Social Philosophy*, 42, 239-53.

Hagen, K. 2010, "Is Infiltration of 'Extremist Groups' Justified?", *International Journal of Applied Philosophy*, 24, 153-68.

Hagen, K. 2011, "Conspiracy Theories and Stylized Facts", *The Journal for Peace and Justice Studies*, 21, 3-22.

Hagen, K. 2018, "Conspiracy Theorists and Monological Beliefs Systems", *Argumenta* 3, 2, 303-26.

Harambam, J. 2017, *The Truth Is Out There: Conspiracy Culture in an Age of Epistemic Instability*, Rotterdam: Erasmus University Rotterdam.

Harris, K. 2018, "What's Epistemically Wrong with Conspiracy Theories", *Royal Institute of Philosophy Supplement*, 84, 235-57.

Hristov, T. 2019, *Impossible Knowledge: Conspiracy Theories, Power, and Truth*, London: Routledge.

Ichino, A. 2018, "Superstitious Confabulations", *Topoi,* 39, 203-17.

Ichino, A. 2019, "Imagination and Belief in Action", *Philosophia,* 47, 5, 1517-34.

Imhof, R. and Lamberry, K. 2017, "Too Special to Be Duped: Need for Uniqueness Motivates Conspiracy Beliefs", *European Journal of Social Psychology*, 47, 724-34.

Jackson, E.G. 2019, "Belief and Credence: Why the Attitude-Type Matters", *Philosophical Studies*, 176, 2477-96.

Kadlac, A. 2015, "The Virtue of Hope", *Ethical Theory and Moral Practice*,18, 337-54.

Keeley, B.L. 1999, "Of Conspiracy Theories", *The Journal of Philosophy*, 96, 109-26.

Knight, P. (ed.), 2002, *Conspiracy Nation: The Politics of Paranoia in Postwar America*, New York: New York University Press.

Levinstein, B. 2019, "Imprecise Epistemic Values and Imprecise Credences", *Australasian Journal of Philosophy* (online).

Lewandowsky, S. *et al*. 2013, "The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science", *Plos One*, 10.

Levy, N. 2007, "Radically Socialized Knowledge and Conspiracy Theories", *Episteme: A Journal of Social Epistemology*, 4, 181-92.

Levy, N. 2019, "Is Conspiracy Theorising Irrational?", *Social Epistemology Review and Reply Collective,* 10, 65-76.

Lyons, J.C. 2009, *Perception and Basic Beliefs. Zombies, Modules, and the Problem of the External World,* Oxford: Oxford University Press.

Martin, A.M. 2011, "Hopes and Dreams", *Philosophy and Phenomenological Research*, 83, 148-73.

Meirav, A. 2009, "The Nature of Hope", *Ratio*, 22, 216-33.

Miceli, M. and Castelfranchi C. 2010, "Hope: The Power of Wish and Possibility", *Theory and Psychology*, 20, 251-76.

Olmsted, K. 2009, *Real Enemies: Conspiracy Theories and American Democracy, World War I to 9/11*, New York: Oxford University Press.

Räikkä, J. and Smilansky, S. 2012, "The Ethics of Alien Attitudes", *The Monist*, 95, 511-32.

Räikkä, J. 2018, "Conspiracies and Conspiracy Theories", *Argumenta*, 3, 2, 205-16.

Recanati, F. 1997, "Can We Believe What We Do Not Understand?", *Mind & Language*, 12, 84-100.

Schultheis, G. 2018, "Living on the Edge: Against Epistemic Permissivism", *Mind*, 127, 863-79.

Sunstein, C.R. and Vermeule, A. 2009, "Conspiracy Theories: Causes and Cures", *The Journal of Political Philosophy*, 17, 202-27.

Sunstein, C.R. 2014, *Conspiracy Theories and Other Dangerous Ideas*, New York: Simon & Schuster.

Swami, V. and Coles R. 2010, "The Truth Is Out There", *The Psychologist*, 23, 560-63.

Swami, V. 2012, "Social Psychological Origins of Conspiracy Theories: The Case of the Jewish Conspiracy Theory in Malaysia", *Frontiers in* Psychology, 3, 280.

Uscinski, J.E. and Parent J.M. 2014, *American Conspiracy Theories*, Oxford: Oxford University Press.

Van Prooijen, J.W. 2012, "Suspicions of Injustice: The Sense-Making Function of Belief in Conspiracy Theories", in Kals, E. and Maes, J. (eds.), *Justice and Conflicts*, Heidelberg: Springer, 121-32.

Van Prooijen, J.-W. and Acker M. 2015, "The Influence of Control on Belief in Conspiracy Theories: Conceptual and Applied Extensions", *Applied Cognitive Psychology*, 29, 753-61.

Van Prooijen, J.-W. 2019, "Belief in Conspiracy Theories", in Forgas, J. and Baumeister, R. (eds.), *The Social Psychology of Gullibility: Conspiracy Theories, Fake News and Irrational Beliefs*, London: Routledge, 319-32.

Velleman, J.D. 2000, "The Aim of Belief", in Id., *The Possibility of Practical Reason*, Oxford: Clarendon Press, 244-81.

Williamson, T. 2000, *Knowledge and Its Limits*, Oxford: Oxford University Press.

Wood, M.J., Douglas, K.M. and Sutton, R.M. 2012, "Dead and Alive: Beliefs in Contradictory Conspiracy Theories", *Social Psychological & Personality Science*, 3, 767-73.

Book Reviews

Dodd, Julian, *Being True to Works of Music*.
Oxford: Oxford University Press, 2020, pp. viii + 194.

The notion of authenticity in musical performance has been discussed extensively in musicology throughout the past century. The story is well-known to those in the music business. Starting from the early 1920s, an ever-growing number of musical historians and practitioners began to engage in the study of pre-classical music, preparing performing editions of ancient works and recreating period instruments. By the 1950s, this new interest consolidated around what is usually referred to as the "Early Music Movement", which contributed significantly to the popularization of the notion of authenticity in the musical field.[1] During the 1980s, however, interest in "authentic performances" began to crack. Musicologists increasingly mistrusted authenticity for being a naïve concept, a misleading ideal, and one giving rise to a series of cold and mechanical performances.[2] In the 1990s, scepticism became so widespread that music scholars gave up talk of "authenticity" altogether.[3]

What makes this story particularly interesting, however, is that right when the Early Music Movement was being administered the final *coup de grâce* by musicologist Richard Taruskin,[4] debates on musical authenticity started to bloom in English-speaking philosophical circles. Burst into the flames of musicology, authenticity—like the legendary phoenix—was born again from its own ashes in the cradle of analytic aesthetics. Since the late 1990s', the release of several important essays by Stephen Davies, Peter Kivy, Roger Scruton, Jerrold Levinson, and many others[5]—all concerned with examining various aspects of performance authenticity and all making explicit usage of the term—marked indeed the emergence of a thriving research field in the philosophy of music, whose ramifications extend to the present day.

There was, in fact, a major twist in the way philosophers of music, as opposed to musicologists, addressed the topic. As Kivy made clear in his seminal volume *Authenticities*,[6] authenticity is not a singular notion, but rather a *plural* one. A context-dependent concept, authenticity remains blurry until we clarify

[1] For a collection of views on the subject of early music see Kenyon, N. (ed.) 1988, *Authenticity and Early Music: A Symposium*, Oxford: Oxford University Press.

[2] See, for example, Taruskin, R. 1984, "The Authenticity Movement Can Become a Positivistic Purgatory, Literalistic and Dehumanizing", *Early Music*, 12, 3-12, "The Pastness of the Present and the Presence of the Past", in Kenyon, N. (ed.), *Authenticity and Early Music*, Oxford: Oxford University Press 1988, 137-210; Dreyfus, L. 1992, "Early Music and the Suppression of the Sublime", *Journal of Musicology*, 10, 114-19.

[3] So much so that it has become usual nowadays to talk about "historically informed" musical performances. See, e.g., Fabian, D. 2001, "The Meaning of Authenticity and The Early Music Movement: A Historical Review", *International Review of the Aesthetics and Sociology of Music,* 32, 2, 153-67.

[4] Taruskin, R. 1995, *Text and Act: Essays on Music and Performance*, Oxford: Oxford University Press, 90-154.

[5] See Davies, S. 2001, *Musical Works and Performances: A Philosophical Exploration*, Oxford: Clarendon Press; Kivy, P. 1995, *Authenticities: Philosophical Reflections on Musical Performance*, Ithaca: Cornell University Press; Scruton, R. 1997, *The Aesthetics of Music*, Oxford: Oxford University Press; Levinson, J. 1990, *Music, Art, and Metaphysics*, Ithaca: Cornell University Press.

[6] Kivy 1995.

Argumenta 7, 1 (2021): 265-277
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.boo

the background against which the term is spelled out, so that there may be not just one but many different *authenticities* in musical performance, all related to, yet all potentially conflicting with one another.

Julian Dodd's latest book *Being True to Works of Music*, published by Oxford University Press in 2020, marks a decisive step forward in the philosophical exploration of the different varieties of authenticity and their consequences on musical practice, as restricted to Western instrumental music. Written in the crystal-clear style of the best analytic philosophy, the author presents readers with a thoroughly-argued examination of the complexities of this controversial notion. Dodd's book, however, offers more than a mere review of the subject. With remarkable insightfulness, it draws an outline of what we might call a "deontology" for music performers, i.e., a sketch of a duty-based approach to decision-making in the field of classical music performance practice. Adopting the ethically-informed vocabulary of obligations, values, conventions, ideals, and norms, Dodd unearths the normative layout and the different value implications underlying musical practice, where "practice" is meant, in a Wittgensteinian framework, as a form of life constituted by a set of specific rules and praxes. While similar normative approaches have already captured attention in other domains of philosophical inquiry—I am referring particularly to recent debates concerning the ethics of reconstructions, restorations, and archaeology[7]—this is still uncharted territory in music. We can only hope, thus, that Dodd's book will be the forerunner of a new season of discussions in this area.

What is it like to be *true* to a work of music? Dodd's provocative title makes us wonder.[8] The answer emerges throughout the six chapters of the book from the close, critical dialogue the author establishes with the major protagonists of this thirty-year debate, and especially Kivy and Davies. Their alternative understandings of authenticity represent the poles against which Dodd elaborates his own proposal. Through a subtle exercise in analysis, Dodd aims to establish an alternative path between the Scylla of Davies' historicist account of authenticity and the Charybdis of Kivy's personalistic approach. As we shall see momentarily, this third path is driven by what Dodd calls "interpretive authenticity", a notion whose articulation constitutes the author's original most contribution to recent debates on the topic.

Chapter 1 of Dodd's book is aimed at providing a theoretical outline of our practice of classical music within which discourses on performance authenticities can be found meaningful. According to Dodd, the most crucial character of our musical practice is that it is intrinsically work-focused, meaning that its artistic endpoint is not to be found in musical performances *simpliciter*, but in how performances are able to present works. An important issue, however, is that scores do not fully determine the sound of accurate performances: they are "gappy" (6). Score translation into sound medium, Dodd argues, "invites interpretation by performers" (5), and as a consequence, the resulting performances

[7] Consider for example Scarre, C. and Scarre, G. 2006, *The Ethics of Archaeology: Philosophical Perspectives on Archaeological Practice*, Cambridge: Cambridge University Press; Bicknell, J., Judkins, J., and Korsmeyer, C. 2019, *Philosophical Perspectives on Ruins, Monuments, and Memorials*, New York: Routledge.

[8] The title is obviously reminiscent of the original meaning of the German term *Werktreue*. For a similar use in the philosophical debate see also Goehr, L. 1989, "Being True to the Work", *The Journal of Aesthetics and Art Criticism*, 47, 1, 1989, 55-67.

can differ even radically from one another, though performers may all be seeking to be faithful to the work. While proving that there can be a plurality of "authenticities", this also means that performers are often left to decide *which* of these versions of "work-faithfulness" deserves priority. The score, as Dodd puts it, can be seen in this sense as a "site of negotiation between composer and performing artist" (155), which implies that conflicts are very likely to arise between the two of them. Most of the text is thus devoted to exploring the possible strategies to resolve these conflicts by scrutinizing and weighing up the normative obligations of each different version of authenticity.

The first form of authenticity Dodd addresses in Chapter 2 is compliance with the composer's written instructions as encoded in the score. Score compliance authenticity (SCA), in Dodd's terms, subtends the idea that performers have a fundamental obligation to maximise *accuracy* when performing musical works. One first problem in this regard is determining whether SCA is a value in our performance practice, i.e., a good-making feature of performance (22). Is playing works just as they were written able to make a performance *ceteris paribus* more aesthetically satisfying than another one? Parting from Davies (2001), who considers SCA as a purely ontological requirement, Dodd offers a positive answer to this question. That SCA is a performance value is demonstrated, he contends, by the fact that listeners *do* believe a performance of a work superior for being more accurate, other things being equal. Accuracy, however, is not just an interpretative option; rather, it is a primary or fundamental goal of our work-focused practice, something that, as he claims, has "final value", meaning that it is valued for its own sake (12).

But what exactly does SCA consist of? Since, as we know already, scores do not completely dictate the sound of any accurate performance, the indications they provide always require interpretation, i.e., need to be disambiguated against the background of some appropriate performance conventions. Which conventions we should adopt to interpret a score becomes thus a substantial question if we aim to achieve accuracy (20). In the philosophical literature, this question has been prevalently treated historically. To comply with the score, it has been argued, performers should read it in light of the musical practices extant at the time of composition, including, but not limited to, the use of period instruments.[9] As Dodd notices, this historicist bent transforms SCA into a form of "historical authenticity" (HA), a notion he addresses in Chapter 3 of the book.

By deploying a vast array of arguments, Dodd's first step is to offer a clearinghouse of all objections that have been raised both in the musicological and in the philosophical domain against HA. He convincingly demonstrates that the problem with HA is not or not so much that it is an unattainable goal, a conceptual naivety, or an indefensible ideal (51). Rather, the problem is that HA does not or cannot do the *normative job* its advocates would like it to do. In other words, while there is a "pro tanto obligation" for performers to respect the score, this does not imply, Dodd argues, that there exists a similar obligation for

[9] Davies, S. 2001, *Musical Works and Performances,* ch. 5; Levinson, J. 1990, *Music, Art, and Metaphysics*; Sharpe, R.A. 1991, "Authenticity Again", *British Journal of Aesthetics* 31, 163-6; Thom, P. 2007, *The Musician as Interpreter*, University Park: Pennsylvania State University.

them to do so by historicist lights. SCA is a final value in our performance practice; HA is not.

How, then, can we articulate SCA so as to escape the historicist lure? Dodd's proposal is that we base score compliance on a more flexible notion of respect for *tradition*, where tradition is understood, in strict opposition to historicism, as a dynamic principle subject to continuous evolution in time with "accumulated wisdom, changing tastes" (63), and new technical developments. On this model, it is the sum of musical practices of the *performer's* time, rather than of the *composer's* (71), that determine the genuine standards of SCA. Although this solution may appear unconvincing to the Early Music nostalgics, it has, it seems to me, the merit of promoting a critical, rather than an antiquarian, view of music history, one in which performers are seen as rooted in a practice that extends back into the past but still admits of innovation; a view where the dead "do not bury the living", to paraphrase Nietzsche.[10]

Having thus rearticulated SCA as a tradition-based norm, Dodd's next move is to confront the major competitor of SCA in musical performance: what Kivy famously calls "personal authenticity" (PA). Chapter 4 of the book is entirely dedicated to the critical investigation of this notion. Following Kivy,[11] Dodd characterises PA as an attempt of the performer to be true to her own *artistic personality* while playing, where "artistic personality" is meant as the sum of the performer's musical tastes, values, commitments, and intuitions. A performance is faithful to the performer's artistic personality in this sense if it bears "the *special stamp*" or is "a *direct extension*" of it (92). These personal marks or imprints, according to Kivy, translate into the way the performance sounds and impinge on its style.[12]

But can PA be also considered a performance value in itself, one resulting in an obligation for performers to shape their performances so that they express their artistic personalities (89)? Guiding the reader through a meticulous dissection of Kivy's thesis, Dodd provides some persuasive arguments against this idea. In Dodd's reconstruction, PA's status as a performance value is grounded in the fact that Kivy considers performances of works as "arrangements" or "versions", thus akin in themselves to "artworks". Since artworks have admittedly greater artistic value the more "personally authentic" they are; and since performances, qua versions, are (akin to) artworks, it follows for Kivy that PA is also a value of performances. Having resumed Kivy's argument in this way, it becomes child game for Dodd to dispute the soundness of the conclusion by denying the second premise. Differently from arrangers, work performers do not seek to produce any new artistic content; therefore, their performances cannot be regarded as artworks (98). This, Dodd clarifies, does not mean that personal style, originality, and creativity do not count as valuable features of a performance, but that such features acquire their value only inasmuch as they contribute to making sense of the work they interpret. Bluntly stated, the fundamental purpose of classical music practice is not that performers express their personality through performance, but that they focus on the work, with an aim to present it to the audience in the most insightful way.

[10] Nietzsche, F. 1983, "On the Uses and Disadvantages of History for Life", in *Untimely Meditations*, Translated by R.J. Hollingdale, New York: Cambridge University Press, 72.
[11] Kivy 1995: 108-42.
[12] Kivy 1995: 123.

This leads us to the last form of authenticity Dodd discusses in the book, which represents, he believes, the most basic norm of performance. The norm is named "interpretive authenticity" (IA) and its analysis is carried out in Chapters 5 and 6 of the book. According to Dodd, IA corresponds to a kind of faithfulness to the work that is associated with, yet distinguished from score compliance. Rather than merely requiring the performer to comply with the composers' written prescriptions, the goal of IA is interrogating the score for the work's meaning, trying to evince the deeper musical content that is embodied in the notes. IA's status as a fundamental performance value, Dodd contends, arises from the fact that works in Western classical music are composed for the sake of being understood, where "understanding" involves grasping their *point* or *integrity* (110) or figuring out their *why* (113). Performers have an obligation to evince and facilitate the audience's understanding of the works they perform. In this sense, complying with the composer's written indications, although necessary, is not sufficient to achieve the subtle understanding of the work's musical meaning that is the aim of musical performance. *Interpretation* is thus essential to present the works insightfully.

In normative terms, this implies that IA, just like SCA, has for Dodd final value, i.e., cannot be traded off for other performance qualities as interestingness, originality, liveliness, and so on. The two authenticities, however, are not on par. Unlike SCA, IA qualifies *as* the "constitutive norm" of our actual practice of classical music, something arising out of what Christine Korsgaard calls the practice's *teleology,*[13] "the raison d'être" (162) of performance practice being to recover the musical meaning expressed in the notes. SCA, by contrast, has its status only as "a conduit to insightful convincing performances" (115). It follows, according to Dodd, that when normative conflicts occur between IA and SCA, our practice approves of the performer's decision to prefer the former to the latter. In other words, disobeying the score's written instructions for the sake of evincing further understanding on the work performed is for Dodd admittable, when done "with appropriately serious-minded and work-focused spirit" (155). Decisions of parting from the score, however, can only be taken according to a conception of what is the right way to present the work and not, *pace* Kivy, for the sake of what "sounds better" (39). The work-focusedness of Western classical music is such that our evaluation of a performance depends on the convincingness of its interpretation of the work. The primary duty of performers is thus to display the work's musical meaning at its best.

It will not escape the reader's notice that Dodd's central argument in support of IA, thus described, rests entirely on a substantive conception of musical meaning in which music is seen as signifying much more than only the notes; an old idea[14] that Dodd, in Chapter 5 of the text, unpacks by adopting Michael Morris' account on the topic.[15] But what that 'much more' stands for is exactly the heart of the issue and represents the Achilles heel of Dodd's thesis. On the

[13] Korsgaard, C. 2008, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford: Oxford University Press; Korsgaard, C. 2009, *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

[14] The *locus classicus* of this idea is Hanslick, E. 2018 (1854), *On the Musically Beautiful*, Translated by L. Rothfarb and C. Landerer, Oxford: Oxford University Press.

[15] Morris, M. 2008, "How Can There Be Works of Art?", *Postgraduate Journal of Aesthetics* 5, 1-18, and Morris, M. 2012, "The Meaning of Music", *The Monist*, 95, 556-86.

one hand, Dodd may be right to insist that musical meaning is never given *a priori* but only revealed by insightful interpretation (129); that it is non-semantic (79), limitless, and resistant to paraphrase (116); and that it is thereby configured, to use Susanne Langer's famous phrasing, as "an unconsummated symbol".[16] On the other hand, however, his argument leaves eventually undecided what this musical meaning is and what it coincides with—whether it is a matter of the pleasantness of the performed sounds, their specific expressive or emotional content, their harmonic structure, or rather the compositional intentions and surrounding cultural and social context of a piece.

Somewhat ecumenically, Dodd's notion of musical meaning holds together all these aspects, leaving a great deal of freedom to both philosophers and performers of music to decide which aspect to privilege and which to sacrifice. While this solution may be accused of circularity, Dodd's insistence on the value of interpretation, it seems to me, has the virtue of reminding us that our musical practice, like all other human practices, is a living thing, always open to revision, and ultimately rooted in the dialectical and stipulative agreement among practitioners.

Roma Tre University LISA GIOMBINI

[16] Langer, S. 1954 (1941), *Philosophy in a New Key: A Study in the Symbolism or Reason, Rite, and Art*, New York: The New American Library, 195.

Earp, Brian D. and Savulescu, Julian, *Love Drugs: The Chemical Future of Relationships*.
Stanford: Stanford University Press, 2020, pp. 280.

In their book, Brian Earp and Julian Savulescu propose a revolution in our way of understanding, living and conceptualizing our love lives: chemistry could be the answer to—many of—our emotional problems, when it comes to falling in and out of love.

Should we take a pill that could ease up our break-up from a toxic—perhaps abusive—relationship? Alternatively, should we not allow for a "chemical boost" in our marriage so to fall again in love with our partner with whom we became unable to communicate after years together?

These are some of the questions that are raised in this unique book. Unique for the angle given to the numerous discussions on love and unique for the academic rigor kept throughout the monograph without losing sight of an accessible and enjoyable prose. The book develops its message through twelve intense chapters that meticulously engage with the main issues at stake when discussing the possibility of "medicalizing love"—from authenticity and social pressure to stigma and mental health.

Concerning the scientific findings in our hands, the book also covers all the available information we currently have on the way our body (and mind?) responds to biochemical inputs from within our body or from outside.

The main thesis of the book is quite simple: we have always (since Romanticism in fact, but we often forget that) described "true love" as one of the highest "goals" we can hope to achieve—hence implicitly depicting it as an unquestionably positive variable in our lives—but that might be misplaced. First of all, love relationships can be extremely unhealthy for us, both physically and psychologically. From toxic and violent relationships to anxiety related to a rejection, many are the instances in which love is *not* the answer. Secondly, from a neurochemical perspective, the very experience of "being in love" has more than one overlapping with situations of addictions to substances such as drugs, alcohol and so on. When in love, we can experience dependency, euphoria and cravings typical of situations of addiction because the dopamine reward system in our brain is activated by our engagement with a romantic partner.

Another "myth" that we tend to associate love with, is that it is "natural"—hence perfect in itself and unchallengeable by definition. The authors want us to rethink this axiom, suggesting that if we were to safely target the neurochemical processes behind romantic attachment that could allow us to help some individuals suffering from different forms of love (lost, rejected, finished), we should—just as we do with other addictions.

Even if all chapters provide plenty of material for discussion and deserve a careful read, possibly the most innovative part of the book (that builds on extensive previous research on the topic by the Earp, Savulescu and colleagues) is the one that focuses on "Ecstasy as Therapy" (Ch. 6), where they put forward a very powerful argument.

Looking into some studies focused on the successful use of MDMA (referred to as "ecstasy" in the street jargon) in cases of Post-Traumatic Stress Disorder (PTSD), Earp and Savulescu challenge us to question some of our *a priori* bias in assessing the medical, social and moral legitimacy of (certain) drugs. If MDMA helps a couple regaining their authentic love through the disempowerment of the defense mechanisms developed by a war veteran when back home to his/her family—that unhealthily damages the possibility to express their true feelings, why should we not use it? People with PTSD might struggle to communicate their love to their partners, resulting closed and unreachable, creating, in time, the conditions for a break-up that would not have taken place had s/he been not blocked in their expressive capacities. MDMA (paralleled with some therapy) might help overcome this block and, in time, dilute it. Using these drugs, the patients (war veterans or otherwise) will be able to lower their guard and let the love of their partners enter their relationship again. The studies are still very limited, but nonetheless deserve a careful evaluation that might shake some of our certainties.

Earp and Savulescu are aware, and make clear, that the chemical dimension of love is not the only factor to be taken into account in the love equation:

> Tinkering with biology, then, is not the only way to modify love: its psychological aspects can be tinkered with as well. At a societal level, people might try to challenge existing narratives about love, including dominant norms on how love should manifest in different relationships. Should love require sex and passion, for example, to count as truly romantic? Or is romantic love more about loyalty and working through difficult problems? Different societies, or the same society over time, might emphasize different factors (22).

This awareness could lead us to very different conclusions—we will not dwell into those here—but what is unquestionably pointed out by the authors is the importance of societal values (that in turn will likely shape our own self-perception) in characterizing and defining our conception of love, but also the way we should relate to it. In other words: why should we have prejudice over the possibility of *preserving* our relationships through the use of some biochemical remodulation?

They write:

> Some people might be concerned that swallowing a pill to achieve insights into one's relationship would be in some sense too easy. The sort of thing that, quickly attained, could just as quickly be lost. But thinking of the pill as an adjunct to relationship therapy should help alleviate this concern. It leaves plenty of room for active, nonsuperficial engagement and intentional learning about oneself and one's partner. As Carol describes the wrap-up to her sessions, the question is, "How are they going to follow up on these insights? They make decisions right there" (90).

Hence, it would appear as if our authenticity would not be undermined by "love drugs" after all, quite the opposite.

Surely Earp and Savulescu's approach to love (and its relationship with more or less legitimate drugs) can be challenged—and it has been done so by a number of authors in a recent special issue of *Philosophy and Public Issues* dedicated to their book.[1] Another criticism that could be moved towards the underlining Posthumanist message that might be extracted (though the authors never frame their argument in those terms)—and structurally rejected[2]—from an embracement of "love enhancing biotechnologies" is that hype for technology that many consider illusionary and promising of too much, unachievable "perfection".[3] Yet, it remains clear that the effort made by the authors in enriching the discussion on this very sensitive (and culturally charged) topic is something to be praised no matter how much we might see it as an "attempted murder" of romantic love.

In conclusion, this is a brave and solid book, written with respect for science and individuals, providing the reader with some innovative ways of looking at and relate to love. Earp and Savulescu's arguments do not need to go unchallenged of course, but engaging with the new, relevant findings that the authors carefully sketch out for us, is something that anyone interested in the topic should really wrestle with—even if critically.

Università LUMSA MIRKO DANIEL GARASIC

[1] Garasic, M.D. (ed.) 2020, "Enhancing Love? Symposium on Brian Earp and Julian Savulescu's book *Love Drugs*", Philosophy and Public Issues, 10, 3.

[2] Levin, S.B. 2021, *Posthuman Bliss? The Failed Promise of Transhumanism*, Oxford: Oxford University Press.

[3] Sandel, M. 2009, *The Case Against Perfection: Ethics in the Age of Genetic Engineering*, Cambridge, MA: Harvard University Press.

Edwards, Douglas, *The Metaphysics of Truth*.
Oxford: Oxford University Press, 2018, pp. x + 198.

One of the main challenges posed by an inquiry into the nature of truth is its apparent connections with several other central and difficult philosophical issues. In recent years, however, the deflationary strand brought promise to dissipate such intricacies and make the notion of truth treatable. Edwards' *The Metaphysics of Truth* is a strong and systematic reaction to this deflationary turn. As Edwards explicitly claims from the beginning, the book has two main purposes. One is that of countering the deflationary and anti-metaphysical approaches to truth; the other is offering a positive, metaphysically loaded conception of what truth is and how it connects to reality. The proposal ultimately consists in a complex form of truth and ontological pluralism. Before entering the book in more detail, it should be remarked that Edwards is a prominent scholar in the field of truth studies, where he distinguished himself as one of the most active and clearest thinkers in the camp of truth pluralism. Indeed, many of the ideas found in the book are already presented and discussed in several published papers. The book is not, however, just the collection of those papers, since Edwards adds many new ideas and combines the various parts in a strong way, from which a well articulated system emerges. The book can be divided into two main parts. The first part includes the first three chapters, and mostly addresses deflationism. From Chapter 4 to 10 Edwards extends the treatment and slowly builds his positive view, with the last two chapters focusing on truth-making theory, and its relations with the pluralist account favoured by Edwards. Let's review the chapters more carefully.

Chapter 1 addresses the question of whether truth should be considered a property or not. Although a positive answer may be considered the most natural and the default option, some philosophers have held different views. On the one hand, there is what Edwards calls the "ultra-deflationist" conception, according to which truth does not define an extension, since it is not expressed by an authentic predicate in the language. On the other hand, truth might be considered an object or an event, rather than a property. The chapter nicely summarizes the main reasons why it is now generally accepted that truth is a property, in at least the basic sense that the truth predicate semantically works as a predicate with an associate extension. The chapter ends with comments on the distinction between concept and property, and on the problem of truth bearers. Willing to avoid endless complications raised by propositions, Edwards adopts sentences as primary truth bearers.

If the truth predicate defines an extension, and, in this sense, it stands for a property, how can deflationists rival traditional conceptions of truth? The standard move is that of pointing out that although it is a property, truth is a very special one. In particular, and against traditional views, deflationists hold that truth is an insubstantial property. But what does this mean? Edwards critically discusses the main options and puts them aside. A widely discussed option, but completely neglected in this book, however, is the clarification of insubstantiality in terms of conservativity. A proposal that sparked a live debate among those working on truth from a formal perspective. Although the metaphysical approach favoured by Edwards is clearly different and distant from a formal perspective, recognition, if not engagement with this other side of the field would have been highly appreciated. In any case, Edwards eventually settles on his own proposal based on Lewis' distinction between sparse and abundant properties. According to him the insubstantiality of deflationary truth is to be understood in terms of abundance. The distinction is one of the key passages in Edwards' strategy and it is crucial in many parts of the book. Roughly, the idea is to distinguish between sparse properties—corresponding to objective similarity and grounding causal-explanatory power—and the abundant ones, which merely correspond to the extension of predicates. Deflationary truth would be an example of the latter, since the truth predicate would define an extension—as the criticisms of ultra-deflationism have shown—but it does not capture an objective feature, as it does not carve reality at any joints.

Once well characterized, a critical discussion of deflationism is offered in Chapter 3. Edwards argues that none of the basic features of the deflationist conception of truth (basicness of Tarskian biconditionals, completeness, purity, insubstantiality) is tenable. Take basicness, which is the feature whose discussion occupies more space. Edwards shows that, far from being basic, as deflationists claim, Tarskian biconditionals can actually be derived and explained in various ways, as done by Tarski, Ramsey, Lynch and Wright. In particular, connections with reference and assertion play a role in deriving disquotational principles, which are thus shown not to be fundamental. Indeed, such connections undermine the other features and put the very insubstantiality of deflationary truth at risk as well. To avoid a complete loss, deflationists must show that such connections do not make truth substantial because also the notions to which truth is connected are not substantial. What emerges is a global deflationism involving a plethora of deflated semantic notions beside truth. This is remarkable. Although deflating other notions usually come natural to proponents of truth deflationism, it goes against the commonly accepted assumption according to which one can be a deflationist about truth without having to be a deflationist about everything else. Also, it is sometimes held that one could easily deflate one notion at the cost of inflating others. By contrast, Edwards' argument shows that a deflationary view only comes as a global package, forcing a deflationary stand over several issues. A consequence of this is that the idea of a "methodological deflationism" is misleading. It relies on the claim that truth deflationism is a neutral, minimally committing view that, as such, should be adopted as the default option and abandoned only if necessary. Deflationism, however, involves highly committing views and is not as innocent as is usually taken to be. If global deflationism is not to be privileged and must defend itself as any other view, then we can have a fresh start and look for a better treatment of semantic notions. This is what Edwards does in the second part of the book, where his positive conception is built.

In Chapter 4 Edwards puts truth aside for a moment. Instead, he investigates how language and the world are connected, focusing in particular on the relation between predicates and properties. By concentrating on the roles that different predicates play, Edwards proposes that predicates come in different kinds. Accordingly, he develops two different models of the relationship between predicates and properties: a responsive model, and a generative model. Roughly speaking, the responsive model corresponds to a realist approach. The idea is that there are sparse properties out there corresponding to objective similarities and causal-explanatory roles, and some predicates just track them. In this sense some predicates respond to sparse properties, namely to mind-independent features of the world. The situation is reversed in the generative model. Edwards argues that some predicates work in an anti-realist way. They do not respond to independent properties, but generate those properties. In this case there are no pre-existing properties that predicates track. Rather, that a certain predicate is satisfied by a certain object is what determines the corresponding property. As a consequence, such a property must be a merely abundant property and not a sparse one. The direction of explanation is now reversed: if "is g" is a generative predicate, an object *t* has the abundant property generated by "is g" *because* the predicate "is g" is satisfied by *t*. Having laid down the models, Edwards discusses some examples. While some of these are not surprising—yet useful to see the approach at work—Edwards interestingly applies the model to institutional, race, and gender predicates. He argues that all such predicates better fit the generative model, and shows how this can help illuminate various philosophical issues. This part is particularly significant for two reasons. One is that thanks to these specific examples, and Edwards' remarkable clarity, good cases for the anti-realist, generative model are offered. This is not obvious, given that the principle "'p' is true *because* p"—rejected in the generative model—is often considered hardly dismissible even by deflationists and anti-realists, unless they are open to embrace an idealistic metaphysics. Secondly, the cases show interesting and new applications of truth theories that are potentially illuminating and important to deal with problems on which traditional debates might appear to have a weaker grip. Note that these merits go beyond the actual correctness of the treatment of the discussed predicates. It does not really matter if Edwards discusses the right conceptions of race and genders predicates in this book. What matters is the kind of role his models play in such contexts.

The chapter just described is key to Edwards' project, since the entire following discussion relies on the distinction between responsive and generative domains. Chapter 5 explicitly extends it to truth. Here the author offers a new argument for a pluralist conception of truth. This is interesting given that truth pluralism is often motivated just by reference to the so-called "scope problem"—according to which traditional conceptions of truth work well in one domain but become problematic when extended to all. The new argument is natural at this point. Edwards shows how the responsive and the generative models involve different forms of truth: a representational (realist) truth and a non representational (anti-realist) truth. By subscribing to the plurality (or at least the duality) of semantic models, one automatically subscribes to pluralism about truth. This chapter also includes an interesting discussion of rival forms of monism. In particular, here Edwards completes his attack against deflationism, showing that (global) deflationism is incoherent. The key step is showing that a deflationary view holding that any property is abundant is incoherent, because abundance requires at least a sparse property of truth. In the next Chapter, 6, an important ingredient is added by extending the pluralist conception to the notion of being. To do so Edwards distinguishes between sparse and abundant objects, where a singular term refers to a sparse object because it exists, and, by contrast, an abundant object exists because its term occurs in a true sentence. Once truth and ontological pluralism are taken on board and combined in a well integrated package, one could wonder how such pluralisms should be understood. The task is completed in the next two chapters, 7 and 8, where the author proposes his version of truth and ontological pluralism respectively. For truth the move is not new, since Edwards rehearses his favourite version of truth pluralism, already defended in other places. Basically, Edwards opts for a moderate form of truth pluralism, according to which there is a single generic truth property, characterized by the usual set of truisms, and a plurality of truth-determining properties, like correspondence and super-assertability. The main idea here is the analogy with the notion of winning a game, which, while general and common to all games, is determined in different ways by different games. The proposal is certainly elegant and clear and Edwards extends it also to existence. An interesting claim of these two chapters is that both truth and existence escape the distinction between sparse and abundant properties. For truth the reason becomes apparent in truth attributions. Consider the sentence "'p' is true". If p belongs to a generative domain, where truth is determined by super-assertability, then, the attribution implies: "p" is superassertible. Since super-assertability is not an abundant property, then "p" is true *because* the sentence to which "p" refers has the sparse property "is true" (the situation is similar for the responsive domain). So, to make sense of truth attributions, truth should apparently be sparse. However, it is typical of abundant properties to be determined by truth, rather than reality. Hence, Edwards concludes that the abundant/sparse distinction does not apply to truth itself.

In Chapters 9 and 10, Edwards discusses the relations between truth-making and truth pluralism. In Chapter 9, the author focuses on an argument (devised by Merricks) for truth primitivism that could be extracted by the claim that some truths do not have a truth maker. Edwards escapes Merricks' argument by leveraging on his pluralism: some truths are not explained in terms of independently existing facts, but they still have a truth maker provided by—the anti-realist notion of—super-assertability. In Chapter 10, the author discusses the argument according to which truth-making theory would make theories of the nature of truth obsolete. Edwards undercuts this strategy by arguing that we cannot understand truth-making without a prior understanding of truth.

Let me now point out some basic objections that could be moved against the strategy presented in the book, before offering a final assessment. For matter of space, I limit myself to two related problems. A first worrying aspect of Edwards' view is the idea that truth escapes the sparse/abundant distinction. This is problematic not just because it is disputable (as I argue next), but because it seems to undermine the very structure of Edwards' maneuver. From the beginning we are told that domains can be classified in different areas, characterized by predicates that stand for either sparse or abundant properties. On this basic distinction both the attack to deflationism and truth pluralism have been motivated. But then we discover that the distinction is not exhaustive: some properties are neither sparse nor abundant. So, what are they? How do the corresponding domains work? And what difference does this make for the attack against deflationism? Can deflationists use that new option? These questions are both crucial and quite natural, but nothing is said in reply. Secondly, one might also propose to read Edwards' arguments for the exceptionality of truth as showing that truth is sparse after all, and thus embracing sparse monism. Indeed, Edwards has not shown that sparse monism is absurd as global deflationism is. Edwards might be dragged to sparse monism also by the treatment of thick predicates (namely moral predicates that have a descriptive content beside an expressive one), offered in Chapter 4. Edwards eventually shows sympathy for the view that all predicates are descriptive and thick to some degrees, even if they are moral predicates. If so, however, a sparse property of truth might be needed to account for the thick ingredient, making the moral domain not merely abundant.

Such objections and potential worries should not prevent us from appreciating the work done. The book is rich with new ideas and applications, and each chapter expands new issues without just repeating itself. The style is very clear and a pleasure to read, the work well structured, comprehensive, and filled with interesting and clever arguments. Such features make it not only an ideal text-book for a graduate course, but also a likely cornerstone of the truth debate to come.

Nanjing University,
Department of Philosophy ANDREA STROLLO

Advisory Board