Charles University

Faculty of Arts

Department of Philosophy and Religious Studies

Philosophy

Disertační práce / PhD Thesis

Martin Zach

Models and Modeling in the Biomedical Sciences

Modely a modelování v biomedicíně

Supervisor: prof. Ladislav Kvasz

Advisor: prof. Roman Frigg

2021

Prohlašuji, že jsem disertační práci napsal samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 4.6.2021 Mgr. Martin Zach, v.r.

Abstract

Many scientific disciplines rely on the construction and use of models: biomedical sciences are no exception. This PhD thesis addresses several aspects of the practice of scientific modeling. First, I discuss the nature of modeling as such, proposing a novel, complementary account of scientific modeling which I term the *experimentation-driven modeling account* and which drives the construction of mechanistic models in many fields of biological and biomedical research, such as cancer immunology. Second, I scrutinize an objection to the mechanistic account of explanation according to which the account fails to accommodate the common practice of idealizing difference-making factors. I argue that this objection ultimately fails because it is riddled with a number of conceptual inconsistencies. Third, I analyze the roles of similarity judgments in some fields of cancer research which employ a variety of mouse models to learn about the disease mechanisms, arguing that by appreciating the epistemic complexities it is possible to shed new light on more general philosophical debates regarding scientific representation. Fourth, mechanisms can also be studied using more theoretical apparatus in the form of simulations. I investigate an example of an agent-based model used to model the outbreak of SARS-CoV-2 and I present reasons for concluding that although these models rely on simplified assumptions, the best of these models can nevertheless be construed as models of actual mechanisms, delivering both mechanistic and difference-making evidence, and serving as tools for evaluating the effects of possible interventions. Finally, I discuss immunology more generally and present a conceptual model of how to think about the immune system. In light of the COVID-19 pandemic and such others as may arise in the future, an adequate understanding of the immune system is required, and philosophy can be of assistance in that regard.

Keywords

abstraction; agent-based models; biomedicine; cancer; COVID-19; idealization; immunology; mechanistic explanation; molecular biology; mouse models; philosophy of science; SARS-CoV-2; scientific models; scientific representation

Abstrakt

Řada vědních oborů, včetně biomedicínských disciplín, vytváří a široce využívá vědecké modely. Hlavní náplní této disertační práce je analýza některých klíčových aspektů praxe vědeckého modelování. Za prvé představím nové, komplementární pojetí vědeckého modelování (modelování založené na experimentování): teprve na této bázi lze adekvátně zachytit praxi vytváření mechanistických modelů v řadě oblastí biologického a biomedicínského výzkumu, včetně nádorové imunologie. Za druhé kriticky prozkoumám jednu z hlavních námitek vůči mechanistickému pojetí vědeckého vysvětlení: jejím cílem je ukázat, že toto pojetí nedokáže uspokojivě zachytit idealizaci rozdílových faktorů, běžně přítomnou ve vědeckém vysvětlení. Vyústěním mé argumentace bude konstatování, že tato námitka zcela selhává vinou řady konceptuálních nesrovnalostí. Za třetí analyzuji různé role, které hrají úsudky o podobnostech v té sféře výzkumu rakoviny, v níž se mechanismy podílející se na rozvoji nádorového onemocnění zkoumají za pomoci různých myších modelů. Pokusím se ukázat, že docenění komplexity těchto úsudků může být významným a zcela původním vkladem do současných filosofických debat o povaze vědecké reprezentace. Za čtvrté poukážu na fakt, že mechanismy lze efektivně zkoumat také prostřednictvím počítačových simulací. Konkrétně se zaměřím na tzv. modely založené na aktérech a jako příklad zvolím model epidemie SARS-CoV-2. Přestože se tyto modely opírají o několik zjednodušujících předpokladů, ty nejpokročilejší z nich lze chápat jako modely reálných mechanismů, poskytující jak mechanistickou evidenci, tak evidenci o rozdílových faktorech: a jak se rovněž pokusím ukázat, mohou se významně uplatnit při vyhodnocování dopadů různých protiepidemických opatření. V závěrečné části práce představím konceptuální model imunitního systému, jehož hlavní zamýšlenou funkcí je přispět ke komplexnímu porozumění povaze imunity. Zásadní význam tohoto úkolu v kontextu aktuální pandemie a možných budoucích ohrožení je naléhavou výzvou i pro současnou filosofii vědy.

Klíčová slova: abstrakce, modely založené na aktérech, biomedicína, rakovina, COVID-19, idealizace, imunologie, mechanistické vysvětlení, molekulární biologie, myší modely, filosofie vědy, SARS-CoV-2, vědecké modely, vědecká reprezentace

List of Contents

Acknowledgements

The completion of this thesis would not have been possible had it not been for the support of a great many people. I would like to thank my supervisor, Ladislav Kvasz, whose engaging and original style sparked my interest in the philosophy of science all those years ago, during my master's program. His continuing support and enthusiasm have helped get me through some difficult periods and provided me with that vital feeling that there is someone upon whom I can rely. I am also extremely grateful to my advisor, Roman Frigg, who invited me to London shortly after we had met, thus providing me with a wonderful opportunity to encounter a much larger world, and which ended up shaping my subsequent research interests. His kindness and feedback, as well as his philosophical rigor, have been truly inspiring.

Additionally, I would like to single out several other people who have left a lasting impact upon my professional development, although it pains me to credit only a few to the detriment of others who also deserve a mention.

The life of a graduate student can occasionally prove extremely stressful due to certain infamous phenomena, such as anxiety and imposter syndrome. I cherish the kindness of Phyllis Illari who, at a 2018 conference in Ghent, helped pick me up when I was down.

Uskali Mäki warmly welcomed me to Helsinki, in so doing greatly helping me to move on with some of my projects.

Words alone cannot express my deepest gratitude to Thomas Pradeu and Maël Lemoine, who first welcomed me to the 'Conceptual Biology & Medicine group' at the ImmunoConcept lab in Bordeaux and, even after my departure, included me in their ongoing work. This has proven essential, especially in light of the COVID-19 pandemic. I owe much to the inspiring environment of the lab in general, and the group in particular. Much of this thesis builds upon the time spent at the immunology lab and on the interactions with the immunologists. I am particularly indebted to Elena Rondeau, Gabriel Marsères, Damien Leleu, Amandine Ferriere, Anne Garreau, Pauline Santa, Julie Déchanet-Merville, Vanja Sisirak, Nicolas Larmonier, Vincent Pitard, and Jean-François Moreau.

Sara Green has been immensely helpful. She provided me with great feedback and advice on various matters.

I feel very much obliged to the East European Network for Philosophy of Science: its inaugural conference, which was held in Sofia (Bulgaria) shortly after I had been accepted onto the PhD program, was my first experience of presenting my work to an international audience. The warm welcome made me feel part of a community. I especially wish to thank Lilia Gurova for her support and counsel, and Daniel Kostic for many memorable moments.

I am also grateful to all my colleagues at the Department of Analytic Philosophy of the Institute of Philosophy of the Czech Academy of Sciences, especially Tomáš Marvan and Juraj Hvorecký, whose guidance and mentorship have been crucial throughout the years.

Additionally, I am truly grateful to several funding bodies: the Hlávka Foundation, the Czech Science Foundation, the Anglo-Czech Educational Fund, the Finnish National Agency for Education (EDUFI), the French government stipend program (BGF), the Faculty of Arts stipend programs, the Charles University Mobility Fund and the Charles University Grant Agency.

Finally, I benefited from numerous interactions and discussions concerning the individual chapters in their different stages of development.

Chapter 1: In particular, I am grateful to Anya Plutynski and Roman Frigg as well as to the members of the Conceptual Biology & Medicine group at the ImmunoConcept lab for detailed feedback. Furthermore, I thank the audiences at the Bordeaux-Sydney Workshop on Philosophy of Biology and Biomedicine (2020) and the online poster session at the Philosophy of Science Association (2021), especially to Lindley Darden and Stephen Downes.

Chapter 2: I am indebted to Roman Frigg and Javier Suárez for providing me with extensive feedback on an earlier version. I thank the audiences at the following events: the research seminar at TINT (Helsinki, 2018), Idealization Across the Sciences workshop (Prague, 2019), ISHPSSB (Oslo, 2019), BSPS (Durham, 2019) and EPSA (Geneva, 2019), where earlier versions of the chapter at various stages of development were presented.

Chapter 3: My work on model representation has undergone substantial changes in response to comments received at the following events: Representation in Science (Prague, 2018), EENPS (Bratislava, 2018), SPSP (Ghent, 2018), Bordeaux group seminar (Bordeaux, 2019), and Scientific Understanding and Representation (Atlanta, 2020). I have also received useful feedback on the chapter from Sara Green, Maël Lemoine, and Roman Frigg.

Chapter 4: I am indebted to Mariusz Maziarz for his collaboration.

Chapter 5: I wish to thank Gregor Greslehner for his collaboration, and for useful feedback I would like to thank the members of the Conceptual Biology & Medicine group at the ImmunoConcept lab and Gérard Eberl from the Pasteur Institute.

Introduction: Models in biomedicine, methods in philosophy

Many scientific disciplines rely on the construction and the use of models – and biomedical sciences are no exception. Scientific literature is notoriously loose when it comes to providing a precise clarification of some of the general concepts such as 'model'. Given that the goals of a scientific paper can be achieved perfectly well without dwelling too much on making the meaning of these general terms more precise, the vagueness should be of no concern. However, since one of the goals of philosophical analysis lies in unpacking such general terms, it must proceed with more care. Furthermore, because modeling is such an essential tool, one may benefit from sharpening the key concepts pertaining to this practice. Indeed, philosophers have long been interested in questions concerning how to characterize the practice of modeling, what models are and how they work. These topics are also the central focus of this thesis. It should be noted that the existing range of topics related to modeling is too vast for any thesis to address and to provide an original contribution rather than a simple restatement or overview of existing views; therefore, a selection had to be made.

Basic biomedical research is, to a large extent, oriented toward studying mechanisms. It is therefore only natural that much of the thesis addresses the question of modeling mechanisms (Chapter 1) or of mechanistic explanation (Chapter 2). To learn about disease mechanisms, biomedical researchers also often rely on the use of animal models such as mouse models (Chapter 3). Mechanisms, however, can also be studied using more theoretical apparatus in the form of simulations, as the case of agent-based modeling of the outbreak of SARS-CoV-2 illustrates (Chapter 4). Although I primarily discuss case studies related to cancer research and cancer immunology, in Chapter 5 I discuss immunology more generally and present a conceptual model of how to think about the immune system. In light of the current pandemic and possible future ones, an adequate understanding of the immune system is called for, and philosophy can be of assistance in that.

More specifically, the following issues will be discussed.

First, I will address the question of how to characterize the practice of modeling. According to the widely held view which I call the *description-driven modeling* account, scientific modeling consists of entertaining a set of model descriptions that specify a model, followed by a detailed investigation of the model and a comparison between the model and the target system. I will argue that this account does not adequately capture important aspects of the practice of mechanistic modeling found in many fields of laboratory research such as cancer immunology. By analyzing research practices concerning the development of mechanistic models of the process of cancer metastasis, I will propose a complementary account which I call the *experimentation-driven modeling* account. On this account, scientists investigate a set of experimental systems and then integrate the results obtained from experiments into a mechanistic model. While the experimentation-driven modeling account shares some key features with the description-driven modeling account, I will argue that the two are epistemically very different research approaches.

Second, mechanistic models are often used to explain phenomena in the biomedical sciences. However, just like most other, if not all, kinds of models, mechanistic models too rely on the use of abstraction and idealization. According to a recent objection, the mechanistic account of explanation fails to account for the common practice of idealizing difference-making factors in models in fields such as molecular biology. I will revisit the debate and I will argue that the objection does not stand up to scrutiny (the work on this chapter gave rise to a paper which is currently under review). This is because it is riddled with a number of conceptual inconsistencies. By attempting to resolve the tensions, I will also draw several general lessons regarding the difficulties of applying abstraction and idealization in scientific practice. Finally, I will argue that more care is needed only when speaking of abstraction and

idealization in a context in which these concepts play an important role in an argument, such as that on mechanistic explanation.

Third, biomedical research also heavily relies on the use of animal models in order to represent various phenomena such as the disease in question. I will focus on mouse models of cancer biology, and in particular cancer immunology, that are used to study tumorigenesis as well as to test cancer therapies in preclinical trials. More specifically, I will devote attention to the immunocompetent and immunodeficient transplantable models, genetically engineered models and humanized models which all exhibit numerous advantages as well as disadvantages. I will then disentangle three – often intertwined in practice but conceptually distinct – research modes: model selection, model extrapolation, and model creation. It will be argued that each of the modes exhibits reliance on different forms of similarity considerations. By appreciating the epistemic complexities, it will be possible to shed some new light on the more general philosophical debates regarding the similarity account of scientific representation. However, rather than to reinvigorate it, this chapter will clarify the specific conditions under which similarity may be crucial for both establishing and maintaining the representational relation between a model and its target.

Fourth, the COVID-19 pandemic caused by SARS-CoV-2 has brought a lot of attention to the epidemiological modeling of the spread of infectious diseases. The outbreak of SARS-CoV-2 required fast-paced decision-making regarding mitigation measures. However, the evidence for the efficacy of non-pharmaceutical interventions such as imposed social distancing and school or workplace closures was scarce: few observational studies use quasi-experimental research designs, and conducting randomized controlled trials seems infeasible. To assess the modeling practice, I will present work done in collaboration with Mariusz Maziarz and published in the *Journal of Evaluation in Clinical Practice* and *History and Philosophy of the Life Sciences*. There, we have considered 'AceMod', an agent-based model (ABM) developed to model the outbreak in Australia. Two points will be argued for. For one thing, although ABMs rely on simplified assumptions, the best ABMs can nevertheless be construed as models of actual mechanisms, delivering both mechanistic and difference-making evidence and serving as tools for evaluating the effects of possible interventions. For another thing, however, there is always the risk that assumptions entertained in ABMs do not include all the key factors and make model predictions susceptible to the problem of confounding. Furthermore, considering that epidemiological ABMs account for not only biological determinants such as infectivity but also social interactions that differ across the globe, the quality of evidence from ABMs must be assessed on a case-by-case basis. In reaching policy decisions, ABMs should be understood as merely one piece of the puzzle subject to further re-evaluation with respect to value judgments. This is because alternative mitigation measures may disproportionately affect certain social groups. Therefore, the quality assessment aimed at identifying possible confounders that have been left out from a particular ABM should delineate the conflict of interest and the vested values related to the ABM and the mitigation measures that it supports.

Fifth, biomedical researchers also propose conceptual frameworks, providing a perspective on the organization of a particular system. Oftentimes they call such a framework a model. Given that efforts like that are conceptual in nature, it is no wonder that philosophers of science have themselves proposed such models. In this chapter I will propose a model of the immune system based on a paper written with Gregor Greslehner and currently under review. In light of the COVID-19 pandemic, there is a pressing need to understand the immune system. We seek to address the mindset from which one views the immune system primarily as a system of defense, which naturally invites the talk of strong immunity reflected in certain areas of immunological research. We argue that although the talk of strong or weak immunity makes sense in a restricted way, such a construal of immunity generally

contributes to the distortion of the overall picture of what the immune system is, what it does, and why it sometimes fails. Instead, we argue for a wider perspective on immunity that is not limited to defense, and we propose a conceptual model to help us understand the immune system in terms of contextuality, regulation, and trade-offs. Thus, reflecting on COVID-19 also allows us to generalize important advances in the understanding of the immune system beyond the current pandemic and health crisis.

The work on this thesis has required the use of various methods of inquiry. In addition to the careful examination of the relevant philosophical literature, I have benefited from pursuing several empirical methods to inform the conceptual analysis (see Zach 2019 for an overview of a variety of such methods). In particular, I have benefited from interviewing scientists and from conducting a participant observation of the research practices during my visit to the ImmunoConcept lab, a research facility located in Bordeaux specializing on research in immunology and cancer.

1. On the nature of scientific modeling: Modeling mechanisms in cancer immunology

1.1. Introduction

In the last several decades philosophers of science have made abundantly clear that much of scientific practice relies on modeling. Indeed, as Axel Gelfert claims, "models (…) are all around us, whether in the natural or social sciences, and any attempt to understand how science works had better account for, and make sense of, this basic fact about scientific practice" (Gelfert 2016, p. v). Philosophers have addressed many perplexing questions concerning scientific modeling. These include issues such as how models explain (Bokulich 2017) and represent (Frigg and Nguyen 2017) phenomena and how they allow for acquiring knowledge about the world in the first place (Frigg and Hartmann 2020; Frigg and Nguyen 2017; Fumagalli 2015; Salis 2016). The question that concerns us here takes us a step back to consider the nature of modeling itself.

According to a widely held view, modeling is an indirect activity of scientific theorizing in which scientists first construct and then investigate a model, rather than the phenomenon itself. I call this view the *description-driven modeling (DDM) account*. In characterizing such modeling practice, some – most notably Weisberg (2007) and Godfrey-Smith (2006) – have distinguished it from a direct strategy of theorizing which they call *abstract direct representation (ADR)*. DDM fits well with much of the scientific practice of modeling. Furthermore, it appears to capture an important sociological or professional dimension of modeling - scientists who are hired and work as modelers.

However, in this chapter I will argue that DDM does not account for the development of mechanistic models in certain branches of biology. Drawing on the method of participant observation and an analysis of the scientific literature, the case of the development of mechanistic models of cancer metastasis, taken from laboratory research on cancer immunology, will be discussed in detail. Consequently, a novel account will be introduced – the *experimentation-driven modeling (EDM) account* – in order to allow for the practices pertaining to mechanistic model building in many fields of biological laboratory research, including cancer immunology. In EDM one derives a model from experiments, that is, one integrates piecemeal experimental results into a unified conceptual framework that is expressed in the form of a mechanistic model, most often in the form of a diagram. I argue that DDM and EDM are distinct modeling practices which nevertheless share several features. Although several differences will be discussed, arguably the most important of these is epistemic: EDM and DDM exhibit different research agendas with respect to modeling. Of note is the point that conflating the two modeling approaches amounts to obscuring important epistemic differences in scientific practices. It is also interesting that scientists involved in EDM are neither sociologically nor professionally recognized as modelers despite the fact that they propose various models.

The structure of this chapter is as follows. Section 1.2. presents the key features of DDM that, according to some of its proponents, set it apart from other ways of theorizing. Section 1.3. provides a primer on cancer immunology and details some of the research practices involved in developing mechanistic models. Section 1.4. introduces and characterizes the EDM account and addresses, in considerable detail, the comparison with DDM and other practices. Section 1.5 elaborates on why EDM should be understood as a *complementary* account.

1.2. The description-driven modeling (DDM) account

The issue of the nature of modeling practices has received relatively little attention compared to many other questions concerning models.[1] Notwithstanding a few exceptions, a consensus about the key characteristics of the modeling process has emerged. The practice of modeling is said to be unified by the representational aspect of models (Giere 1988; Glennan 2017; Hughes 1997; Teller 2001). In Teller's words, "in principle, anything can be a model, and (…) what makes a thing a model is the fact that it is regarded or used as a representation of something by the model users" (Teller 2001, p. 397). In particular, modeling is characterized as a practice in which scientists represent the target systems *indirectly*: they engage in an indirect theoretical investigation and draw surrogative inferences (Contessa 2007; Frigg and Nguyen 2017; Giere 1988; Godfrey-Smith 2006; Knuuttila 2017; Knuuttila and Loettgers 2017; Levy and Currie 2015; Mäki 2009, 2011; Morrison 2015; Morrison and Morgan 1999; Parkkinen 2017; Salis 2016, 2019; Thomson-Jones 2020; Weber 2014; Weisberg 2007, 2013). The indirectness of modeling will become clear as soon as we look closely at the stages of the modeling process. According to Weisberg (2007), there are roughly three such stages, which can be described as follows:

1. *Model construction*. In the first stage, scientists construct a model by means of entertaining certain model descriptions.[2]
2. *Model analysis*. In the second step, the properties and the dynamics of the model are investigated.
3. *Model comparison*. Finally, the model is assessed by comparing the model outcomes with its target.[3]

It must be noted, however, that according to Weisberg (2013, p. 74) the stages of modeling - while conceptually distinct – do not necessarily take place in this rigid order as they may happen together or iteratively. Still, it does seem safe to assume that in order for scientists to study a model, some version of a model must first be proposed.

Similarly, Godfrey-Smith (2006) describes the modeling process as consisting of the specification and investigation of a hypothetical system, i.e., a model, followed by the consideration of resemblance relations between the hypothetical and the real-world systems (see also, e.g., Frigg 2010; Salis 2016).

[1] For instance, Morrison and Morgan (1999, pp. 12–13) noted that "we are given definitions of models, but remarkably few accounts of how they are constructed." Since then, the situation has improved, although much has remained the same. This is because a large portion of the philosophy of modeling has focused on the nature of models, i.e., the ontological question, rather than on the nature of modeling as a practice. While some have explicitly drawn a distinction between the two, others have not. For instance, Weisberg (2007, p. 208) admits that "there are many insightful discussions in the philosophical literature about the nature of models" but that "less has been written explicitly about the practice of theorizing." On the other hand, Toon (2010) speaks of "the ontology of theoretical modelling," and makes further remarks which may be viewed as collapsing the distinction, at least to some extent (see also, e.g., Thomson-Jones 2012).

[2] Model descriptions are taken to be assumptions, equations, parameters, pictures, empirical data, words or pieces of text or any such 'thing' that give rise to models or model systems, whatever the ontological status of these may be (Frigg and Nguyen 2017; Godfrey-Smith 2006; Mäki 2009; Thomasson 2020; Thomson-Jones 2010; Weisberg 2013). It should be noted that some authors have questioned some aspects of the distinction between model descriptions and models (Knuuttila 2017; Odenbaugh 2015).

[3] Note that this third step is optional. Although it is true that such comparison often takes place, sometimes models are constructed and investigated independently of any real-world phenomenon against which they could be compared (Mäki 2009; Thomson-Jones 2020; Weisberg 2004, 2007, 2013).

Thus, scientists construct models as stand-ins for complex phenomena, and instead of investigating the target phenomena directly, they investigate them indirectly, that is, take a detour through the model.

To illustrate this general schema, Weisberg (2007, 2013) introduces the Lotka-Volterra model of predator-prey dynamics.[4] Because the activities of the fishing industry during World War I dropped off significantly, one would intuitively expect there to have been an abundance of fish after the war ended. However, once the war was over it turned out that there was a shortage of various kinds of fish in the Adriatic. Surprisingly, it was observed that the population of sharks and rays seemed to have increased, while the population of squid and cod had decreased. To understand why this was so, Lotka and Volterra, independently of one another, constructed a system of two coupled differential equations describing the hypothetical populations of predators and prey. In particular, the equations describe how the population dynamics are coupled. Following the construction and analysis of the model, Volterra figured out that whereas low levels of general biocide – which kills both predators and prey – would provide favorable circumstances for population growth in predators, high levels would contribute to population growth in prey.

Seen in this way, Weisberg argues that Volterra first constructed a model using certain model descriptions expressed in the form of mathematical equations (i.e., model construction). He then analyzed the model by studying its dynamics (i.e., model analysis). Finally, the qualitative predications were matched against the available data (i.e., model comparison).

I call this view the *description-driven modeling (DDM) account* because the modeling practice proceeds by entertaining certain model descriptions, on the basis of which a model is constructed and then investigated instead of investigating the target system directly.[5]

This indirect strategy is not the only one that scientists have at their disposal. Both Weisberg (2007) and Godfrey-Smith (2006) speak of another approach to scientific theorizing, distinct from DDM, called *abstract direct representation (ADR)*.[6] In contrast to DDM, scientists engaged in ADR represent and analyze phenomena without the mediation of a model, i.e., they investigate the phenomenon directly. As Godfrey-Smith (2006, p. 734) puts it, "one approach is to immediately try to identify and describe the actual system's parts and their workings. A distinct approach is to deliberately describe another system, a simpler hypothetical system, and try to understand that other system's workings first."

An example of the ADR practice provided by Godfrey-Smith concerns a book by Leo Buss from 1987 called *The Evolution of Individuality*, which is contrasted with the DDM approach exhibited by the 1995

[4] For a more detailed exposition see Weisberg (2013, pp. 10–13) and especially Knuuttila and Loettgers (2017) who offer a critical and more historically-oriented description.

[5] I believe that the term *description-driven modeling* can serve as an umbrella term for a number of modeling strategies already discussed in the literature. This includes, among others, both the *theory-driven modeling* strategy, in which modeling is regulated by general theories, and the *phenomenological modeling* strategy, in which semi-empirical results and concepts beyond the theory framework are used (see Portides 2011). It also includes *autonomous modeling*, where models are developed independently of a strongly empirically-confirmed framework theory (see Reutlinger et al. 2018). In cases where data mining practices lead to the construction of network models (see Plutynski and Bertolaso 2018), much of data-driven modeling can be also viewed as an instance of DDM. The overall modeling process mirrors the steps characteristic of DDM: the construction of a network model followed by the analysis of the features of the network (and the comparison with the phenomenon).

[6] Note that while the great majority of authors think of modeling as an indirect activity to be distinguished from a direct, non-modeling way of doing science, there are a few authors who disagree and argue for a direct view of modeling (see Levy 2012, 2015; Toon 2012a).

book, *The Major Transitions in Evolution*, by Maynard-Smith and Száthmary. Both books address the question of the origin of multi-cellularity. However, whereas Maynard-Smith and Száthmary rely on the modeling approach, Buss' work is model-free and consists of a detailed examination and careful analysis of "the actual relations between cellular reproduction and whole-organism reproduction in known organisms" (Godfrey-Smith 2006, p. 731). Godfrey-Smith argues that Buss' arguments, while cautious and speculative at times, are based on the causal roles and the consequences of actual cellular machineries, their environmental circumstances, and the developmental sequences, rather than on a deliberate consideration of simplified or otherwise schematic organisms. Similarly, Weisberg (2007) discusses the work of Mendeleev as an illustration of ADR practice. According to Weisberg, by examining the properties of chemical elements, Mendeleev created a representational system that captured a pattern exhibited by the chemical elements. Thus, in contrast to indirectly representing the phenomenon by creating and studying a model, as Volterra did, Mendeleev's approach was direct in that he represented trends in chemical reactivity rather than trends in a model system.[7]

Although Weisberg and Godfrey-Smith are in agreement with regard to the general distinction between DDM and ADR,[8] they diverge on some specific issues. Weisberg (2007, p. 228) admits that "it may be possible to take the equations that describe Volterra's model and treat them as approximate, direct representations of Adriatic predator and prey populations." However, he further claims that the fact that "these transformations may be possible should not change our analysis of their theoretical practice" because "the contrast between modeling and ADR is about the practice, not the products of theorizing" (Weisberg 2007, p. 228). Godfrey-Smith (2006, p. 734) appears to be somewhat more liberal, warning us that "it would be a mistake to say that the distinction is always so easy to draw" and that there are "unresolved problems to tackle in this area." At the same time, he suggests that there is a sociological dimension to modeling, something which will be addressed in more detail in Section 1.5.

Thus, modeling – according to the DDM account – is an indirect strategy of scientific theorizing whereby scientists first construct a model by entertaining certain model descriptions and later devote much effort to its analysis.

1.3. Cancer immunology

Let us now turn to a brief overview and conceptual introduction to the field of cancer immunology, followed by a discussion of experimental practices involved in the study of the role of the myeloid-derived suppressor cells in cancer metastasis. This case study provides important lessons which will prove crucial to the introduction of the novel account of scientific modeling expounded in Section 1.4.

1.3.1. A primer on cancer immunology

Cancer has been a major topic of biomedical research for well over a century, but only recently has it caught the attention of philosophers.[9] The definition of cancer alone presents problems: it has been

[7] It should be noted that some authors have questioned Weisberg's analysis of Mendeleev's work, as well as the strict distinction between direct and indirect approaches (see Knuuttila and Loettgers 2017, p. 1012 for a discussion and a list of references).

[8] Both agree on the indirectness of modeling and the stages in which modeling happens. They are also keen to stress that although abstraction, idealization and other tools are part and parcel of the modeling process, they are not unique to modeling (see especially Weisberg 2007, pp. 228–229).

[9] See, for instance, the monographs by Bertolaso (2016), Laplane (2016), or Plutynski (2018). See also a recent entry in the Stanford Encyclopedia of Philosophy (Plutynski 2019).

variously defined as (i) a disease of cells, (ii) a developmental disease, or (iii) a disease of a tissue (Weinberg 2014). As Robert Weinberg, a prominent cancer biologist notes, gene-centric and cell-autonomous views have been abandoned by the mainstream in favor of some sort of anti-reductionistic picture according to which numerous factors beside the tumor cells themselves play an essential role in the process of tumorigenesis.

Cancer immunosurveillance – the idea (developed by Paul Ehrlich and later elaborated by Sir Frank MacFarlane Burnet) that the immune system plays an active role in keeping most tumors in check by seeking them out and destroying them – has long been questioned to the point of having virtually been abandoned (Davies 2018 Chapter 8; see, e.g., Decker et al. 2017; Dobosz and Dzieciątkowski 2019; Pradeu 2019). The evidence against the immunosurveillance hypothesis has come from diverse sources. According to mid-20th century cutting edge science, tumor cells appeared phenotypically so similar to healthy cells that the immune system was believed to be tolerant of the tumors. The experimental evidence came from studies conducted on mouse models in which a state of immunodeficiency was experimentally induced by neonatal thymectomy (Dunn et al. 2002). Although the results were mixed, there appeared to be no difference in the incidence of chemically-induced tumors between the immunodeficient and the wild-type mice.[10] The crucial experiments providing what seemed to be the decisive evidence took place in the 1960s and 1970s when Stutman (1974) used *nude* mice, a strain that is naturally missing thymus due to a mutation in a single gene. Here again, the chemical induction of tumors led to no observed difference in the tumor burden of both the immunodeficient and wild-type mouse.

However, since then further evidence has been amassed and the idea of immunosurveillance has been reinvigorated. It has turned out that the immunodeficient mouse strains used in the above experiments were not as immunodeficient as initially thought (Dunn et al. 2002).[11] In contrast, mice with mutated genes coding for RAG proteins do not develop a functional adaptive immune system, making them severely immunodeficient. Studies showed that RAG-deficient mice experience an increased tumor incidence even in tumors of non-viral origin; additionally, epidemiological evidence from transplant medicine provided support to the immunosurveillance hypothesis (Dunn et al. 2002). Some transplant patients later developed cancer. Transplant recipients take immunosuppressive drugs to prevent their immune system rejecting the donated organ. A careful examination of the available data showed that in these cases, the donors suffered from and were later proclaimed to have been cured from cancer years before donating their organs. To make sense of these and other findings, the notion of immunoediting (Dunn et al. 2002) proves extremely helpful. The process of immunoediting consists of three phases: (i) the elimination phase (i.e., immunosurveillance), during which the immune system destroys tumor cells; (ii) the equilibrium phase, in which variants of the tumor cells that are poorly immunogenic or that have acquired ways of subverting the immune system evade destruction; and (iii) the escape phase, which marks the point at which tumor cells are no longer being kept at bay.

One of the reasons why tumor cells can evade immune destruction and later escape is that the specific environment in which tumor cells arise, called the tumor microenvironment, is generally

[10] To be more precise, there was a difference with respect to the incidence of tumors of viral origin. However, such an observation could be expected, given that the immunodeficiency resulting from removing the thymus leads to a decrease in the ability to control viral infections. The reasoning was as follows: had the immune system been implicated in controlling tumors, there should have been an observed increase in the incidence of tumors of non-viral origins too.

[11] The impression was that mice lacking thymus would have no functional T cells. It turned out that although limited in numbers, these mice have fully functional classical αβT cells, γδ T cells, and NKT cells, all of which play a role in immunosurveillance.

immunosuppressive (Weinberg 2014).[12] By releasing various substances with immunosuppressive potential and by actively recruiting immune cells while inducing a suppressive phenotype in them, cancer cells create a milieu which allows them to escape surveillance. Thus, the immune system actually plays a paradoxical, dual role in cancer: it eliminates tumors but may also promote tumor growth.

The immune system is implicated not only in tumorigenesis but also in the metastatic process. An intriguing observation regarding metastasis, the cause of death in 90% of all cancers, is its apparent tropism, i.e., a tumor arising in a particular tissue is likely to metastasize to a set of particular organs but not to others: for instance, it has been well established that breast cancer tends to metastasize into lungs, bone, brain and liver (Weinberg 2014). Noting this surprising phenomenon, Stephen Paget, a 19th century British surgeon and pathologist, proposed the "seed and soil" hypothesis, arguing that a tumor (the seed) can only grow if it lands on a fertile ground (the soil).[13]

Current research indicates that the metastatic organs undergo changes before the arrival of cancer cells (Liu and Cao 2016). Thus, rather than being a passive recipient of the "seed", the "soil" is actively being transformed in a complex dynamic process that gives rise to a pre-metastatic niche which ultimately leads to the establishment of a secondary tumor site (Liu and Cao 2016). One of the key players implicated in establishing a pre-metastatic niche are myeloid-derived suppressor cells (MDSCs), a heterogeneous population of immature cells of myeloid origin activated under pathological conditions (Gabrilovich and Nagaraj 2009).[14] It is common to distinguish two broad sets of these cells based on different expression patterns, namely the monocytic and granulocytic MDSCs (M-MDSCs and G-MDSCs respectively), with the occasional mention of a third, early-stage population of MDSCs (Veglia et al. 2018).[15]

1.3.2. Experimental inquiry into the role of myeloid-derived suppressor cells in cancer metastasis

Inquiry into the role of MDSCs is an ongoing process with still many unknowns. Such research projects rely heavily upon studying experimental systems such as cell cultures and animal models and make use of a vast array of experimental assays. In what follows, I provide a brief – and in no way exhaustive – description of some of the common methods used in cancer immunology, based in part on the use of participant observation method in an immunology lab. In particular, much of the discussion concerns

[12] Philosophers have contributed to a key conceptual debate regarding the nature and boundaries of the tumor microenvironment (Laplane et al. 2018; Laplane, Duluc, et al. 2019).

[13] This has sparked a debate in the community and multiple competing theories have been proposed to account for the observed metastatic tropism (see Fidler 2003). According to some (e.g., Weinberg 2014), the seed-and-soil hypothesis is promising, even though it may fail to explain certain features of the metastasis such as the rarity of contralateral metastases (i.e. tumor cells disseminated from, for example, one breast should be naturally seeded to the other breast which should provide the most hospitable environment).

[14] As discussed by, for example, Veglia *et al.* (2018), MDSCs have been found to play a biological role not only in cancer but also in infectious diseases, autoimmunity disorders, obesity, and pregnancy. Although potency and the particular mechanisms by which the subsets of MDSCs mediate their immunosuppressive effects vary depending on the site (e.g., lymph node / tumor microenvironment), they do so in both an antigen-specific and nonspecific manner, and they suppress the adaptive as well as the innate immune system (Kumar et al. 2016; Nagaraj et al. 2010).

[15] As is often the case when defining new cell subpopulations, some have recently questioned whether current state-of-the-art knowledge warrants the talk of MDSCs as a category of cells distinct from monocytes and neutrophils of a particular phenotype (see Garner and de Visser 2020, BOX 4).

the project of Elena Rondeau, in which the goal is to provide the characterization of MDSCs in different organs and in different time points in metastatic breast cancer in order to gain insight into the exact role played in that context by MDSCs. Please note that this particular example should be understood as exemplifying the practice found across many different (biological and other) fields in which much of the focus is devoted to wet lab research, allowing us to draw general lessons and to formulate a philosophical account of the practices involved (see Section 1.4.).

Using cell cultures in experimental practice often involves some preparatory work, such as the use of a lentiviral vector in order to introduce the genes for the enzyme luciferase and the green fluorescent protein (GFP) in the 4T1 cell line, which is a standardized breast cancer cell line. Beyond that, various functional assays are conducted using co-culture experiments. These include studying the immunosuppressive effects of MDSCs on healthy T cells, or the migration behavior and changes in phenotype when co-culturing MDSCs and tumor cells.

Animal models also play an essential role in studies of cancer metastasis. The BALB/c mouse strain serves as the recipient of the 4T1 cell line modified by the lentiviral vector, which results in a cohort of tumor-bearing mice models. Mice of the same strain also serve as controls and as a reservoir of healthy T cells that may be used in co-culture experiments. Organs are taken from both the tumor-bearing mice and the healthy controls and resected into tissue slices of approximately one cell layer (10-12 µm). These tissue slices are then subjected to immunohistochemical investigation, the goal of which is to search for and locate both the MDSCs and the metastases in lungs. Specific antibodies are used to stain the MDSCs, thus allowing for visualization. This process is repeated at different stages of tumor development to provide further data, e.g., if there are any changes over time in the number, position, and type of MDSCs.

Visualization methods are also crucial when conducting *in vivo* experiments. By day 11 after injecting the immunocompetent BALB/c mice with 4T1 breast cancer cells, the mice exhibit tumors of approximately 7-10 mm. They are then injected with luciferin, a substrate that binds the luciferase enzyme expressed by the 4T1 cells, resulting in bioluminescence. This allows for the localizing of tumor cells in a living animal; and by day 20, metastases start to appear. The mice are then killed and dissected, and their organs investigated using the same imaging technique, thus providing additional precision. Imaging methods allow for important observations, yet they fail to provide important insight into the mechanisms responsible for the observations. To that end, cells are collected from organs and subjected to further analysis using a variety of experimental instruments. Among the essential lab equipment is the flow cytometer. A sample of cells suspended in a fluid and often labeled with fluorescent markers is injected into the flow cytometer, and flowing one at a time, the cells pass through a laser. Scattered light is then detected and processed by a computer. In short, flow cytometry is a method that enables the detection and measurement of some physical and chemical properties of cells. Fluorescence-activated cell sorting (FACS), a feature of many of the flow cytometers, allows for the gathering of cells of a particular type for later analysis. For instance, the M-MDSCs and G-MDSCs can be detected and sorted on the basis of their expression of CD11b and CD45 markers and distinguished from one another by a difference in their level of expression of Ly6C and Ly6G markers.[16] Selected for their surface markers, these cells can then be subjected to a polymerase chain reaction (PCR) to analyze the gene expression patterns: taken from different organs at different times, these

[16] M-MDSCs are commonly characterized as $CD11b^{+}$ $Ly6G^{-}$ $Ly6C^{high}$ cells, whereas G-MDSCs as $CD11b^{+}$ $Ly6G^{+}$ $Ly6C^{low}$. Note that this is valid only for mice cells because human cells do not express Gr1 – thus, no Ly6C or Ly6G epitopes.

cells are probed to reveal the set of factors that may be characteristic of them at different sites and at different times, such factors including, among others, cytokines and tissue-specific chemokines.

Additionally, a lot of research makes use of a variety of excitatory and inhibitory studies (see Craver and Darden 2013 for an extended discussion). For instance, studies have focused on investigating the impact of depleting MDSCs on the formation of metastasis (Ouzounova et al. 2017). Likewise, knock-out experiments are conducted with the same goal in mind: for instance, one can knock out a gene coding for a chemokine such as CXCR2 which has been implicated in recruiting MDSCs to the tumor site to see what the effect will be (Katoh et al. 2013). The results show a decrease in MDSC recruitment to the secondary site and a resulting decrease in metastatic tumor burden; conversely, transferring wild type MDSCs to $CXCR2^{-}$ mice leads to an increased metastatic burden (Katoh et al. 2013).

Combining all these and other methods, scientists can begin to generate results which help them develop an overall picture of the metastatic process.

1.4. Introducing the experimentation-driven modeling (EDM) account

Having provided some background context and detailed some of the experimental methods used in cancer immunology research, we may now formulate some philosophical conclusions. Cancer immunologists seek to discover and understand the processes by which primary tumors metastasize. They provide accounts of phenomena which they often express by means of diagrams, which in turn are taken to represent mechanisms (see Figure 1.1). Indeed, the scientific literature is full of references to mechanisms and mechanistic models. Mechanisms and their models have been a hot topic in the philosophy of science literature for the past two decades. An ecumenical view has emerged regarding the minimal characterization of mechanisms, according to which "a mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon" (Illari and Williamson 2012, p. 123). There are two points of note. First, what scientists often express by a diagram is a representation of a purported mechanism – a mechanistic model. Second, mechanistic models are models of phenomena.[17] In our case, the target phenomenon is a complex process, namely the formation of a pre-metastatic niche.[18]

It is important to note that the way in which cancer immunologists proceed in developing mechanistic models is to a great extent different from the modeling process described by the DDM account. In order to provide a more accurate description of modeling in various fields of biological research, another account must be proposed.

[17] Models of phenomena have long been distinguished from a set of interrelated notions such as data models or models of experiments (Bogen and Woodward 1988; Giere 2010; La Caze 2011; Leonelli 2019; Woodward 2011). Whereas the former notion concerns models of the phenomena of interest, the latter pertains to finding patterns in the data and the use of statistical and other data processing methods.

[18] Note that here a model of a phenomenon accounts for a process rather than an entity.

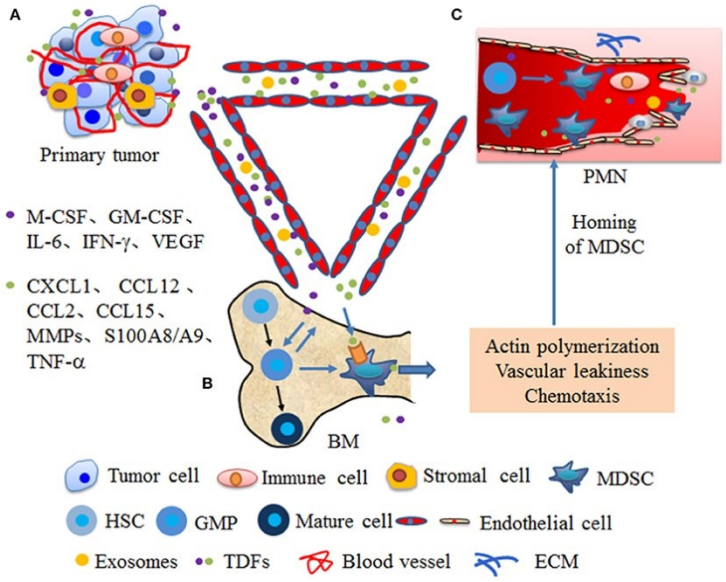


Figure 1.1. An example of a mechanistic model expressed in the form of a diagram. Figure adopted from Wang et al. (2019).

1.4.1. What is EDM?

I argue that the process of building mechanistic models in fields such as cancer immunology can best be captured by what I call the *experimentation-driven modeling* (EDM) account.[19] The EDM account is best defined as *the practice of integrating piecemeal experimental results into a comprehensive conceptual framework which is expressed in the form of a mechanistic model*.[20]

In EDM, scientific investigation starts with choosing a set of experimental systems that are relatively easy to control and manipulate, an example of which is a cell culture or an animal model. A variety of the experiments described above are routinely conducted on these systems, leading to the production of experimental results which are taken to provide some (albeit limited) insight into the nature of the studied phenomenon. When sufficient experimental results have been produced, a more complete picture begins to form, ultimately giving rise to a mechanistic model introduced to account for, at least to some extent, the target phenomenon. This is a long and piecemeal process which often requires combining results from a multitude of studies, including those published by other teams. Once a

[19] The terms *theory-driven models* and *experiment-driven models* appear, for instance, in the work of Mitchell and Gronenborn (2017), who discuss modeling approaches in the context of research on protein structures. Although similar to some extent, the way I discuss modeling in this paper differs from their approach. For Mitchell and Gronenborn, whereas theory-driven modeling concerns the practice of predicting the protein structure by means of running computations from physical and chemical principles, experiment-driven modeling pertains to algorithmically inferring models from experimental data. However, the notions of DDM and EDM as presented here are more general: DDM covers not only theory-driven approaches but also theory-independent approaches, and EDM also covers approaches that are much less, if at all, guided by any sort of algorithm and as such the derivation seems much less straightforward.

[20] Fagan (2016) outlines a similar position while discussing research on human embryonic stem cells. In the context of immunology, Baetu (2014) has claimed that the "big picture" of some pathway or mechanism of interest is often built up as a mosaic of scientific knowledge (see also Lemoine 2017 for similar remarks). Mitchell and Gronenborn (2017) have discussed a variety of both theoretical and experimental approaches to modeling the structure of proteins, the ways in which these approaches may be integrated and how they complement one another.

mechanistic model is proposed, an additional process of equal rigor is required to further validate the model experimentally and to figure out its possible limitations. This is because, despite the fact that the models are derived from experimental results, the experimental systems are in many ways artificial and are subject to generating results that provide distorted pictures of what is going on in the actual full-blown phenomenon. By way of example, let us consider a set of experiments done in the 1990s concerning research on graft rejection. It was known that the presence of specific CD8 T cells is crucial for rejecting grafts. Knocking out the perforin gene in mice – one of the mechanisms by which CD8 T cells kill cells – generated T cells which lack the ability to kill graft cells *in vitro*. However, there was no observed difference in rejecting skin grafts between the perforin-less and the wild type mice (Clark 2007, p. 218). What works *in vitro* may work differently, or not at all, *in vivo*, and vice versa.

Furthermore, because experimental results are often highly sensitive to a particular experimental context, it is not uncommon to find data that limit the extent to which the mechanistic model can be applied to account for what would intuitively be considered as one and the same phenomenon. Going back to cancer immunology, consider the following example. Ouzounova *et al.* (2017) report that whereas M-MDSCs switch on the epithelial-mesenchymal transition (EMT), thus facilitating the dissemination of tumor cells, G-MDSCs act to change the phenotype of disseminated cells through the process of mesenchymal-epithelial transition (MET), allowing for the establishment of a micro-metastasis. In contrast, referring to the study of Toh *et al.* (2011), Condamine *et al.* (2015) state that G-MDSCs rather than M-MDSCs are responsible for initiating EMT. One of the many possible explanations for this discrepancy lies in the fact that while Ouzounova *et al.* studied the 4T1-induced breast cancer mouse model, Toh *et al.*'s findings concern a mouse model of spontaneous melanoma. Thus, model comparison in EDM may also not be completely straightforward.

As previously noted, models produced by EDM are derived from experimental results. However, because EDM accounts for mechanistic modeling, one should be careful not to conflate EDM with models that have been referred to by a variety of terms such as descriptive, phenomenological, or black-box models, which merely summarize data without committing to underlying mechanisms (Craver 2006; Glennan 2017; Kaplan 2011).

Finally, the steps in EDM can be thus rendered explicit:

1. *Analysis of experimental systems*. Experimental systems are manipulated to generate experimental results.
2. *Model construction*. Experimental results serve as building blocks in the construction of a mechanistic model which accounts for the studied phenomenon.
3. *Model comparison*. To estimate the extent to which the model adequately accounts for its target, scientists often conduct further experiments in addition to changing the experimental context.

1.4.2. On the differences between EDM, DDM, and ADR

With these three steps now explicit, the comparison between DDM and EDM should become apparent. However, two caveats must be considered.

First, the particular steps in which mechanistic models in cancer immunology are derived from experiments appear to differ from the steps in DDM. Recall that in DDM, the modeling process happens in roughly three stages: scientists first use model descriptions to construct a model as a stand-in for the target phenomenon (*model construction*); they then investigate the model to find out what it

implies (*model analysis*); and finally, they compare the model results with the target phenomenon (*model comparison*). Thus, one of the differences is that whereas DDM begins by constructing a model, followed by its analysis, EDM's starting point is an experimental investigation which ultimately leads to model construction. However, as previously stated, Weisberg (2013, p. 74) notes that the stages of DDM do not necessarily occur in this rigid order as they may happen together or iteratively. Still, as argued, the order of steps does seem to be representative of much of the practice, setting it apart from EDM.

Second, one may also wonder to what extent EDM and DDM are, in fact, distinct as clearly both can rely on experimental results. Recall that model descriptions that give rise to models as per the DDM approach can be not only assumptions but also empirical data, among other things. Clearly, then, the line between the two cannot be drawn on such terms. However, a closer inspection reveals an important difference not to be missed. First, EDM engages in the laborious processes of experimental data generation whereas DDM more often relies upon pre-existing data. Thus, the kinds of expertise required are often very different. Second – and more importantly – the crucial difference lies in the crux of the research practices involved in the two modeling approaches: while the crux of the work in DDM is the study of the model, in EDM the work is basically considered done once a model is proposed. In other words, DDM is best characterized by "playing around" with a given model, and although models also serve cognitive purposes in EDM, e.g., to provide a comprehensive picture of the mechanism, EDM does not "play around" with models. Conflating the two modeling approaches would thus obscure important epistemic differences in scientific practices.

The proponents of DDM are very keen to stress the indirectness of the modeling practice, which sets it apart from other ways of doing science, such as ADR. The question remains to what extent EDM satisfies the requirement of indirectness, a crucial feature of modeling according to the proponents of the DDM account. Should it turn out that EDM lacks this key feature, perhaps it ought not to be considered a modeling practice but either an instance of ADR or yet another, significantly different way of doing science. I argue that there are good reasons for maintaining the claim that EDM does indeed possess the feature of indirectness. The fact remains that in EDM the mechanistic model is not the central focus of scientists as it is in DDM. Thus, the purported indirectness of modeling according to the EDM account cannot stem from the same source as in DDM. However, neither in EDM does one study the phenomenon directly: the focus of investigation is a set of experimental systems that are assumed to capture – often in a highly artificial way – some aspects of the phenomenon. Thus, in EDM scientists investigate the phenomenon indirectly via a detour through the investigation of a set of simpler systems. Therefore, EDM does exhibit the feature of indirectness.

Perhaps it is less controversial to claim that a mechanistic model is the result of what I call the EDM practice here, than it is to claim that the practice is, in fact, a modeling practice. In describing the key features that distinguish modeling from ADR, Weisberg warns us about conflating the process leading up to the product with the product itself. The practices of DDM and ADR are to be

> "distinguished by the actions and intentions of theorists, not by the outcome of the process of theorizing. This means that to judge whether or not a particular theorist is a modeler, it will not be sufficient to determine whether or not her theory can be represented as a model or cluster of models. We will actually need to know something about how the theory was developed and how the modeler set about trying to represent the world" (Weisberg 2007).

Weisberg further claims that

> "modeling is distinguished from ADR by a theorist's construction and analysis of a model, which is used to analyze and represent a real-world phenomenon indirectly if at all. When a modeler wants to describe a real phenomenon, she begins by choosing a model, not a real phenomenon to analyze" (Weisberg 2007)

and that in ADR

> "the theorist is analyzing a representation that is directly related to a real phenomenon, anything she discovers in her analysis of the representation is a discovery about the phenomenon itself, assuming that it was represented properly. There is no extra stage where the theorist must coordinate the model to a real phenomenon" (Weisberg 2007).

However, as argued above, cancer immunologists ordinarily choose a set of experimental systems as the focus of investigation in order ultimately to learn something about the target phenomenon. They do not directly analyze the phenomenon. Rather, they intentionally pursue the indirect line of investigation and because research conducted on cell cultures and animal models does not straightforwardly translate to knowledge about the target phenomenon (see also below), there is an extra stage at which the scientist must coordinate her results with the real phenomenon. Thus, EDM differs from ADR.

In addition to comparing EDM with DDM in terms of the intentions and steps by which these practices proceed, one can also turn the spotlight onto the role of assumptions in both kinds of approaches. In DDM, assumptions serve as a kind of model description and are said to be the building blocks of models: by entertaining certain assumptions, scientists construct models. Often this is put in the following terms: scientists write down model descriptions by means of which they create model systems (Frigg and Nguyen 2017; Godfrey-Smith 2006; Mäki 2009; Thomasson 2020; Thomson-Jones 2010; Weisberg 2013). For instance, Volterra wrote down equations (i.e., the model descriptions) describing the relations between two hypothesized populations (i.e., the model system).

In EDM, the role of assumptions is quite different. They neither define nor otherwise give rise to models. Instead of serving as building blocks for creating model systems, assumptions in EDM concern the representativeness of the experimental systems and the validity of experimental results with respect to the studied phenomenon.[21] They also help in deciding which experimental systems to use in the study of a particular phenomenon. As noted above, much of cancer immunology research makes use of cancer cell lines such as the 4T1. Although these standardized cell lines originate from tumor biopsies, it is well known that once they have been adapted to cell culture conditions, they no longer behave like the tumors arising spontaneously *in vivo*. The genetic and/or epigenetic changes in these cells lead to cell immortalization, meaning that they can proliferate indefinitely. Cancer cell lines grow in an environment without the need for heterotypic interactions – a type of communication between different cell types which controls the proliferation of the other types of cells in the neighborhood – which sets them apart from the tumors originally found in cancer patients.[22] The two-dimensional spatial arrangement of cell cultures and other features also introduce conditions not found *in vivo*.[23]

[21] See also Weber (2014) who argues that the role of assumptions in using model organisms concerns things such as the validity of the results and as such is different from the role of assumptions in mathematical modeling.

[22] To be more precise, some carcinoma cells *in vivo* develop to the state where they no longer depend on stromal support and can grow and proliferate independently (Weinberg 2014).

[23] Although many labs now routinely use three-dimensional cultures, known as organoids or spheroids, they comprise only a small part of cell culture research.

Recall the example of perforin-less CD8 T cells: under the *in vitro* conditions they lose their ability to reject grafts, yet they seem to do their job perfectly fine *in vivo*.

The research is further complicated by a well-known feature of biological systems, namely redundancy. It is often the case that while the inhibition of a particular pathway may seem promising in one experimental context, it ultimately leads to disappointing results in another because a back-up pathway takes over.

Some might wonder whether the key difference between EDM and DDM cannot be drawn along the ontological dimension. Philosophers have long distinguished physical or material models from theoretical or nonmaterial models (Frigg and Nguyen 2016; see, e.g., Toon 2010). Because EDM seems to concern material practices as opposed to the theoretical practices of DDM, perhaps EDM should be understood as an account of material modeling, whereas DDM could be an account of theoretical modeling. This view is mistaken for two reasons. Firstly, in EDM, material practices are involved mainly in step 1; step 2 is theoretical in that it concerns not the manipulation of material systems but the integration of piecemeal experimental results into a mechanistic model.[24] Moreover, the mechanistic model is not a material entity in the straightforward sense; rather, it is a conceptual model expressed in the form of a diagram. Secondly, many material models are clear cases of DDM, e.g., the San Francisco Delta-Bay model (Weisberg 2013) and the Phillips-Newlyn machine, a material model of macroeconomics (Frigg and Nguyen 2018). These physical models are constructed as simplified versions of their target systems and are subsequently investigated in order to learn about the features of their respective targets.[25]

It would also be wrong to draw a line between EDM and DDM in terms of exploratory versus hypothesis-driven research because there are plenty of examples of both the exploratory and the hypothesis-driven instances of DDM and the same can be said of EDM (Gelfert 2016, Chapter 4). In practice, there may also be instances in which the distinction may not be that sharp, and a researcher working on a project may also be moving back-and-forth between the two extremes.

[24] At this point it is worth noting the difference between the EDM account and some of the other accounts that discuss experimental modeling, such as the experimental modeling account of Weber (2014). Weber claims that experimental modeling "consists of constructing model systems that are composed of living organisms (sometimes, but not necessarily, genetically modified) and that are used as *in vivo* representations of biological processes in such a way that some processes are used as stand-ins for other processes" (Weber 2014, p. 787). This, then, differs from the EDM account in two important respects. First, Weber's experimental modeling basically concerns only the first step in EDM: it is the investigation of experimental systems. Second, when Weber speaks of constructing model systems, he means the construction of experimental systems; in EDM the construction of models pertains to constructing conceptual mechanistic models (step 2).

[25] Depending on the exact context of the research, it should also be noted that there might be cases in which it is not perfectly clear whether modeling should be thought of in terms of EDM or DDM. For example, in one of the previous footnotes the experimental modeling of the structure of proteins is discussed. Frigg and Nguyen (2016) discuss Kendrew's model of myoglobin, a material model of a protein, along the lines of DDM. In their own words, although the "model was constructed on the basis of electron density data (…) it wasn't simply a summary of these data, or a tool to communicate effectively the information the data contained. The model provided epistemic access to the tertiary structure of the molecule in a way that the electron density data alone could not" (Frigg and Nguyen 2016, p. 226). Frigg and Nguyen claim that some of the key insights regarding myoglobin came from studying its material model. Thus, at least in this case it seems natural to construe the work as consisting of constructing a model, later followed by its analysis – a picture that fits DDM. However, in other cases it might be more natural to think of the experimental modeling of proteins in terms of EDM (see Mitchell and Gronenborn 2017 for a discussion on modeling proteins). What this seems to highlight is the difference in research goals. One way or another, it is possible that the difficulty with classification may be more pertinent to models of entities than to models of mechanisms.

Finally, EDM is somewhat peculiar in that the process of mechanistic modeling builds on the use of other types of models, namely the animal models and cell cultures which are also sometimes referred to as models, a point also clearly articulated in Fagan (2016). Arguably, while in DDM scientists sometimes also proceed by constructing a model on the basis of another model, there does not appear to be such a built-in dependency as there is in EDM. This is because while all cases of EDM exhibit the modeling hierarchy, there are clear-cut examples of DDM in which the construction proceeds without the mediation of another model or theory.[26]

1.4.3. Experimenting and modeling

Because experimentation is part and parcel of the EDM account, one could gain the false impression that any instance of experimentation amounts to modeling. One can run an experiment or even a series of experiments without piecing the results together into a mechanistic model. It is only when there is an effort to understand the mechanism responsible for the phenomenon of interest by running a series of experiments, the results of which are ultimately accounted for by developing a model, that we can speak of EDM. Experimentation on its own should not be conflated with modeling; indeed, although for different reasons, this point has already been made in the existing literature (see, e.g., Weber 2014).

That said, the philosophical literature brings forth a number of interesting analogies between modeling and experimenting on the basis of which it concludes that experimenting *is* modeling (see, e.g., Mäki 2005).[27] One important analogy suggested by Mäki pertains to his use of the notion of isolation, the act of removing influences deemed, at least provisionally, irrelevant to the task at hand.[28] In the case of theoretical modeling, this means employing assumptions that neutralize the influence of disturbing factors, whereas in material modeling the experimental systems are isolated in a lab and sheltered from the causal influences of the outer world. Thus, both the theoretical and the material manipulations are viewed as isolations. However, as noted before, the simple fact that abstractions, idealizations – or, in this case, isolations – are typical of modeling does not warrant the conclusion that other practices that make use of isolations should therefore be equated with modeling. For present purposes, however, there is no need to take a firm stance on this issue: what matters is that the EDM account does not rest on equating experimenting with modeling. Rather, the EDM account views modeling as the practice of integrating results obtained from experiments conducted on a set of experimental systems – material models, if you will – into a mechanistic, that is to say, a conceptual model. None of this would require commitment to the claim that experimenting, *per se*, is modeling.

[26] For instance, think of autonomous models such as Schelling's model of social segregation (Reutlinger et al. 2018).

[27] Mäki is careful not to commit to the strong reading of this thesis, i.e., the claim that *any* modeling should count as experimenting, and *vice versa*. In his own words: "The equation models=experiments is not suggested to hold for all specifications of the two concepts, that of model and that of experiments. The equation rather boils down to two more specific claims: many theoretical models=experiments, and many material experiments=models" (Mäki 2005, p. 312). However, since he gives no specific examples of cases in which this analogy breaks down, I take the liberty of discussing this issue with respect to EDM.

[28] Mäki (2005) lists additional analogies. However, addressing them is beyond the scope of this paper.

1.4.4. EDM: The bottom line

Before moving on, let us summarize the key aspects of EDM. It is the practice of integrating piecemeal experimental results into a comprehensive conceptual framework which is expressed in the form of a mechanistic model, often as a diagram. In EDM, scientists first choose a set of experimental systems that, albeit in a distorted fashion, are assumed to capture some of the salient features exhibited by the target phenomenon. By running experiments, scientists produce data. The results are then used to construct a mechanistic model. Rather than investigating or otherwise playing around with the model, scientists work with a set of laboratory experimental systems: the mechanistic model is then the end product of the modeling process. Finally, the extent to which the mechanistic model correctly accounts for the phenomenon of interest may be assessed on the basis of additional experimental verification.

Because there is no direct investigation of the target phenomenon, EDM comprises an indirect analysis (akin to DDM, as opposed to ADR). Although EDM is not assumption-free, the role of assumptions concerns the representativeness of the experimental systems and the validity of results. In contrast to DDM, the assumptions do not give rise to models. The difference between EDM and DDM is not captured by either of the distinctions between material and non-material models, and exploratory and hypothesis-driven models, respectively. Crucially, EDM rests on model hierarchy: mechanistic models are constructed on the basis of investigating a set of experimental (model) systems. Finally, EDM does not equate experimenting with modeling.

1.5. EDM as a complementary account of scientific modeling

The EDM account of scientific modeling should be understood as complementing rather than replacing or modifying the widely held DDM account. This is because while the way in which DDM is characterized accounts for a large portion of research practice in various fields, it does not fit well with those practices employed in modeling mechanisms in many fields of biological research, such as cancer immunology.

1.5.1. Between the normative and the descriptive approaches

To fully defend EDM as a complementary account there are at least two problems which must be addressed. The first concerns the sociological dimension of modeling; the other pertains to the role of philosophical analysis. Let us address them in order.

One may question the extent to which it is justified to speak of modeling in the context of mechanistic models based on laboratory research practices. After all, the practice of modeling seems to exhibit a sociological dimension in that “some scientists now are trained, hired, and assessed as modelers; that is their job description” and that “modelers have their own subculture within science, to some extent, and their own language” (Godfrey-Smith 2006, pp. 728–729). Indeed, seen in this way it would be difficult to maintain the claim that many cancer immunologists working in experimental labs engage in modeling, for the simple reason that they are not hired as modelers and they do not perceive themselves as such.

There is no doubt that the sociological aspect sits well with the DDM account: it is likely that most, if not all scientists who work within the DDM framework are in fact hired and assessed as modelers. In contrast, scientists whose work is best captured by the EDM account are hardly ever hired as modelers. Nevertheless, there are other (epistemically) important features shared by EDM and DDM: when

building mechanistic models of phenomena, EDM scientists proceed indirectly: instead of directly analyzing the target phenomenon, they construct and investigate a set of simpler systems on the basis of which they construct conceptual mechanistic models. Thus, while the sociological dimension is a unique feature of DDM, it does not prevent us from construing the EDM practices as modeling practices.

Scientific literature is notoriously loose when it comes to providing a precise clarification of some general concepts, such as a 'model'. Given that the goals of a scientific paper can be achieved perfectly well without dwelling too much on making the meaning of these general terms more precise, the vagueness should be of no concern. However, since one of the goals of philosophical analysis lies in unpacking such general terms, it must proceed with more care. Roughly, two extreme views can be discerned in this context. Whilst philosophical analysis might espouse a strictly descriptive approach and consider anything referred to by the term 'model' as an instance of a model, constructed by some modeling practices, one can stipulate the meaning of a given term by providing a philosophical analysis of a set of presumably paradigmatic examples while excluding possible alternatives. This latter approach exhibits strong normative tendencies. The EDM account, much like that of DDM, is situated somewhere between these extremes. On the one hand, it takes seriously the notion of a mechanistic model; yet on the other hand, it has built-in boundaries specified by the key characteristics described in the previous section. Similarly, according to some of its proponents, the DDM account should be understood as providing an incomplete picture of modeling practice.

Along those lines, Weisberg argues that "just as theorists offer incomplete, idealized models of their targets, so must philosophers. Theoretical practice is rich and multilayered, and the world is often uncooperative" (Weisberg 2013, p. 6), to which he further adds that "by developing philosophical accounts of modeling, we can start to get a handle on theoretical practice. But just as in a representation of any other complex phenomenon, philosophical analysis will necessarily be partial and incomplete. Thus the accounts developed in this book are themselves models of modeling" (Weisberg 2013, p. 6). The EDM account helps to partially complete the picture by providing another piece of the puzzle.

1.6. Chapter summary

Scientific modeling is an important tool in contemporary science. Philosophers of science have long discussed many aspects of the practice of modeling. My review of the description-driven modeling account, and my proposal of the experimentation-driven modeling account, demonstrate firstly, that DDM does not account for the practice of mechanistic modeling in laboratory fields such as cancer immunology; secondly, that EDM fits well with what is going on in cancer immunology research and beyond; and thirdly, that EDM should be understood as a complementary account which can coexist with DDM.

2. Modeling via abstraction and idealization: How not to criticize mechanistic explanation

2.1. Introduction

The world in which we live is immensely complex. Indeed, its complexity vastly exceeds our capacity to grasp it in its entirety with all the exact detail in place. Nevertheless, scientists do more than a decent job of keeping the complexity in check by constructing models of selected phenomena that help us to understand, explain, predict and control various aspects of the world. To achieve this, models must be simple enough to facilitate insight into the phenomena. In the literature of the past several decades, much has been said about the nature(s) and function(s) of models (see also Chapter 1).[29] Many, though not all, authors prefer to speak of abstraction and idealization as examples of tools for introducing simplifications into models. Overall, models are commonly considered to be relatively poor in detail and often to provide distorted accounts of their target systems.

Importantly, abstractions and idealizations have also been discussed in the context of specific philosophical debates such as that on the mechanistic account of explanation. In a recent paper, Alan Love and Marco Nathan (2015) have argued that the new mechanists' preferred view of explanation cannot account for the common practice of idealizing difference-making factors in models in molecular biology.

This chapter scrutinizes the analysis provided by Love and Nathan and argues against their conclusion that the mechanistic account of explanation is in trouble. More specifically, I will argue that their analysis paints a confusing picture for a number of reasons: it is interwoven with inconsistencies regarding (i) how they treat one and the same modeling assumption, (ii) how they apply their preferred definitions, and (iii) how they formulate the core objection. Moreover, the assumptions that they present as idealizations can – instead, and perhaps more naturally – be accounted for in terms of abstractions. In the process, I also draw several general conclusions for the debate on abstraction and idealization and its use. For one thing, it will be shown that philosophers developing accounts of these notions often disagree among themselves with respect to a number of issues, meaning that the notions might not be as clear cut as generally believed. Relatedly, while the distinction between abstraction and idealization is relatively easy to spell out, it proves extremely tricky to adequately apply it in scientific practice. This may, in part, be due to the fact that the various existing accounts have been developed in different disciplinary contexts; and applying the distinction originally developed in one context to another may not be a straightforward process, for it may overlook important differences in epistemic practices characteristic of the respective disciplines. Finally, the arguments laid out in this chapter should also serve as a cautionary note to those who have embraced the objection to the mechanists, not realizing the fundamental issues underlying such criticism. More generally, philosophers may need to pay special attention when using the concepts of abstraction and idealization before these concepts can do any real work in a philosophical argument.

The structure of this chapter is as follow. Section 2.2. discusses abstraction and idealization and notes that while some papers present detailed accounts of these notions, other papers rely on a ready-made

[29] See, e.g., Frigg and Hartmann (2020) for a comprehensive review. Extended discussions can also be found in a number of monographs or book editions. (Bailer-Jones 2009; Cartwright 1983; Gelfert 2016; Giere 1988; Magnani and Bertolotti 2017; Morgan 2012; Morgan and Morrison 1999; Morrison 2015; Toon 2012a; Weisberg 2013)

distinction. Section 2.3. reviews the key parts of Love and Nathan's analysis, including their case study on models of gene expression, which is followed by a detailed critical examination in Section 2.4. Section 2.5. provides concluding remarks.

2.2. Abstraction and idealization

The philosophical debates over the last several decades have made abundantly clear that much scientific practice relies upon the construction of models that, in some sense, are simplified versions of their target systems. Among the most frequently mentioned techniques for introducing simplifying assumptions into models are the practices of abstraction and idealization. It should be noted that these two, however, hardly present an exhaustive list of the forms of simplification. Indeed, many other notions are commonly discussed, some with partially overlapping or otherwise interconnected meanings. Furthermore, I should note that the term 'assumption' is somewhat ambiguous. As suggested in Chapter 1, assumptions may either be viewed as the building blocks of models (as in DDM), or they may concern the representativeness of the experimental systems and the validity of experimental results (as in EDM). However, throughout this chapter I will use the term 'assumption' indiscriminately to refer to any kind of an assumption involved in modeling.

An important distinction concerns the nature of these assumptions on the one hand, and on the other hand, the roles these assumptions play. For instance, Demetris Portides states clearly that the "character [of idealization] and its epistemological implications" (Portides 2013, p. 253) are separate issues. Philosophers usually tend to focus exclusively on one or the other, perhaps somewhat more rarely on both (e.g., Potochnik 2017).[30] This is, of course, a perfectly legitimate endeavor which reflects various interests.[31] However, as should become clear in Section 2.4., what is at stake here is both the former issue, i.e. the character of abstraction and idealization, and the question of what implications it has for the latter, i.e. for the way in which these concepts figure in at least some explanatory practices in molecular biology.

One set of papers that deals with the topic of abstraction and idealization often relies on a ready-made distinction. In the simplest terms, idealization is construed in terms of (deliberate) distortion, misrepresentation and/or falsehood and amounts to providing an inaccurate picture of the studied system, whereas abstraction concerns the omission of an (irrelevant) feature.[32]

---

[30] For example, when presenting the notions of Galilean idealization, minimalist idealization and multiple-models idealization, Michael Weisberg claims that "despite the differences between minimalist idealization and Galilean idealization, minimalist idealizers could in principle produce an identical model to Galilean idealizers" and that "the most important differences between Galilean and minimalist idealization are the ways that they are justified. Even when they produce the same representations, they can be distinguished by the rationales they give for idealization" (Weisberg 2013, p. 102). Arguably, then, Weisberg's Galilean and minimalist idealizations are (or at least can be) one and the same (kind of) assumption that is put to work in different ways.

[31] Much focus has been devoted to inquiring into those various functions. For instance, they may allow for making a model mathematically tractable (Jebeile 2017), although it should be noted that it is not necessarily the case that all assumptions are limited to mathematical modeling. Abstraction and idealization can also help in isolating difference-making factors by narrowing down the focus of a model (Mäki 1992; Strevens 2008; Weisberg 2013), and they play various roles in explanation (Batterman 2009; Bokulich 2011; Jebeile and Kennedy 2015; Kennedy 2012; Reiss 2012; Rice 2015; Rohwer and Rice 2013; Wayne 2011) or understanding (Elgin 2007, 2017; Potochnik 2015, 2017; Reutlinger et al. 2018; Rice 2016; Strevens 2017).

[32] Commenting on the available relevant literature, Margaret Morrison states that "most of [the literature] draws a distinction between idealization which is construed as the distortion of a particular property (e.g.

For instance, according to Roman Frigg and Stephan Hartmann,

> "Aristotelian idealization [i.e., abstraction] amounts to 'stripping away', in our imagination, all properties from a concrete object that we believe are not relevant to the problem at hand" (Frigg and Hartmann 2020).

> "Galilean idealizations [i.e., idealization] are ones that involve deliberate distortions" (Frigg and Hartmann 2020).

Arnon Levy and William Bechtel claim that,

> "Broadly understood, idealization is the introduction into a theoretical model of simplifying falsehoods—assumptions that are known not to describe accurately the target phenomenon but that nevertheless expedite analysis and understanding. To say that a population of rabbits is infinitely large is an idealization, in this sense. Insofar as a model is abstract, it need not contain any falsehood or inaccuracy. Abstractions are poor in detail yet potentially true and accurate. Idealizations are by definition mismatched to reality" (Levy and Bechtel 2013, p. 243).

While referring to Godfrey-Smith (2009), Alan Love and Marco Nathan state that abstraction concerns "the intentional omission of detail" and that "abstraction must be distinguished from idealization, the deliberate misrepresentation of detail in a model" (Love and Nathan 2015, p. 763). Similarly, in the words of Marta Halina, models "are abstract in the sense of omitting detail about the target system and idealized in the sense of distorting elements of that system" (Halina 2018, p. 219).

Referring to Jones (2005), Mazviita Chirimuuta explains that "by 'abstract' [she means] a model which leaves out much biophysical detail, in other words 'highly incomplete'; by 'idealized' [she means] a model which describes a system in an inaccurate or unrealistic way" (Chirimuuta 2014, p. 133).

Finally, Worth Boone and Gualtiero Piccinini state that

> "mathematical and computational models are typically constructed not only by abstracting away from many details of the target system but also by replacing those details with simplifications and idealizations that distort or misrepresent the target system" (Boone and Piccinini 2016, p. 680).

As hinted above, some of the authors who use the distinction explicitly draw on another set of papers which aims to provide a more nuanced characterization of the terms by clarifying in what precise sense idealizations may be thought of in terms of distortion, misrepresentation and falsehood, and abstraction in terms of omission of details (e.g., Godfrey-Smith 2009; M. R. Jones 2005; Levy 2018; Mäki 1992; Portides 2018; Potochnik 2017). Although there exists a consensus among the authors developing more nuanced accounts on some of the general features such as the need to distinguish between at least two meanings of the process of omitting certain features in a model – one of which may better be understood as an idealization rather than an abstraction – many other issues remain a matter of debate. For example, while some authors including Levy (2018) argue that idealization must be intentional, such requirement is explicitly denied in Jones (2005). Among many other things, philosophers also disagree on the question whether idealizations are best construed as concerning

frictionless planes) and abstraction which involves the omission of properties (e.g. a body's material in calculating its trajectory)" (Morrison 2011, p. 343).

individual claims (Levy 2018), or whether they should be thought of holistically, that is, as applicable to models as wholes (Rice 2018, 2019).[33, 34]

Importantly, authors developing accounts of abstraction and idealization often characterize the notions against a backdrop of a specific disciplinary context, ranging from economics to various

[33] Another interesting issue concerns, for instance, the extent to which it is legitimate to construe abstraction in terms of omission-as-subtraction, a term coined by Portides (2018). On this view, according to Portides, abstraction amounts to stripping away – subtracting – features which presupposes that the modeler knows beforehand whether or not the modeled system in fact possesses the features in question. Portides claims that the omission-as-subtraction view is implicit in how abstraction is commonly understood, and he cites Cartwright's work to illustrate the point. Cartwright states that abstraction "is not a matter of *changing* any particular features or properties, but rather of *subtracting*" (Cartwright 1989, p. 187) and that "abstraction in science works by subtracting all those factors which are only locally relevant to the effect" (Cartwright 1989, p. 224). This interpretation is suggestive and is often motivated by a handful of examples. For instance, when modeling the movement of a ball by writing down the equations of motion, color is subtracted from the description since it is considered irrelevant to the task at hand (Cartwright 1989, p. 187). Similarly, Julie Jebeile states that "there is no point in specifying the Moon phase for describing the motion of a body, or, of including the presence of oxygen or the average temperature for describing the trajectory of planets" and that "the description, once cleared of the less relevant details, contains all the relevant representational aspects" (Jebeile 2017, p. 216). Although abstraction by subtracting features is indeed descriptive of cases in which enough knowledge has accumulated, i.e., where scientists are fairly familiar with many of the features of the studied system, it may be more adequate to characterize abstraction generally in terms of omission-as-extraction, whereby certain features are 'extracted' from the system irrespective of what the other features may be (Portides 2018). This is because, although omission-as-subtraction may adequately capture what is going on in domains in which the concept was developed, the fact remains that such characterization fails as a general characterization of abstraction – a thing to note when applying it to fields other than those in which the notion was originally discussed (Portides presents an example from quantum mechanics to that effect). Indeed, in many areas including much of molecular biology the details are being filled in in a piecemeal fashion rather than crossed out from the outset. For instance, prior to figuring out the detailed workings of a signaling pathway, the exact molecular complexes involved in the pathway are usually not known (see Craver and Darden 2013). Additionally, and in parallel to explaining how models are built, the philosophical literature has also addressed the ontological question of what models are. Abstraction as subtraction naturally fits with some metaphysical debates: the process of subtracting features generates abstract objects and scientific models have been construed as such (Giere 1988; Glennan 2017; Mäki 2009; Psillos 2011; Teller 2001). Thus, some of the accounts could be interpreted as dealing with the question of ontology for which it is presumably well equipped (but see Frigg and Nguyen 2017; Thomson-Jones 2010; Toon 2012a for arguments against models as abstract objects). However, some (N. Jones 2018; Levy 2013) have explicitly warned against making connections between abstraction employed in the service of constructing scientific representations and abstract objects.

[34] In fact, there is a host of other issues that philosophers have addressed and that are directly relevant to elaborating the concepts further. Among these issues we find questions related to the nature and function(s) of abstraction and idealization, many of which remain a subject of controversy. For instance, consider the following questions: how exactly do abstraction and idealization relate to truth (M. R. Jones 2005; Levy 2018; Portides 2018; Teller 2012), to mathematics (Jebeile 2017), to fictions (Bokulich 2011; Suárez 2009), or to approximations (M. R. Jones 2005; Morrison 1998; Norton 2012; Portides 2007)? How should we adjust our views regarding the historical development of theories? More specifically, can we reinterpret certain theories or models from the past *as if* such theories were postulating simplifying assumptions even though these 'assumptions' used to be taken at face value? Or should we refrain from such practice and instead consider only an intentional usage as possible instances of these assumptions? Are they eliminable, should they always be, and how are they justified (Batterman 2002, 2009; Batterman and Rice 2014; Bokulich 2017)? Are they best construed as concerning individual claims or rather holistically as pertaining to the models as wholes (M. R. Jones 2005; Levy 2018; Rice 2019)? Do they come in degrees? And if so, how can we estimate different degrees (M. R. Jones 2005; Levy 2018)? Should we think of them in terms of the processes by which models are built or as model products? All these issues, many of which remain unresolved, suggest that the topic of abstraction and idealization is, in fact, very complex.

domains of physics and of biology. Extrapolating the distinction from one disciplinary context and applying it in another may prove challenging (see Section 2.4.1.), for the range of practices in these disciplines may differ considerably.

Naturally, some of the authors who develop accounts of abstraction and idealization then end up using them, or *vice versa*. For instance, Levy in Levy and Bechtel (2013) quoted above simply uses the distinction, referring to Jones (2005) and Godfrey-Smith (2009), while in his (2018) he develops a more nuanced account.[35]

Finally, one may wonder whether a more thorough conceptual analysis of abstraction and idealization is required. In fact, two reasons may be offered for thinking it has both practical and otherwise important consequences (see also Levy 2018). Providing conceptual clarity with respect to the notions discussed here may prove useful for scientists who engage in various methodological debates in their community, including the arguments over the issues of the *realisticness* of assumptions in providing understanding of phenomena (Mäki 1992). The other reason, the one that concerns us here, is to avoid using arguments in philosophical discussions which are grounded in concepts that are ill-defined for the purposes at hand. Failing to clarify the key concepts may generate great misunderstandings and lead to cycles of fruitless debates, generating even more confusion.

2.3. Love and Nathan on the mechanistic account of explanation in molecular biology

In a recent paper, Alan Love and Marco Nathan (2015) present a case against the mechanistic account of explanation, arguing that the account fails as it is unable to account for what they see as the widespread practice of idealizing difference-making factors in scientific models, which, thus, brings us back to the debate on abstraction and idealization. In particular, they argue that "the intentional misrepresentation of causal relations, which are the source of explanatory power in a description of a mechanism's components and activities, generates a significant—albeit neglected—problem for the mechanistic framework" (Love and Nathan 2015, p. 762).

There are several key concepts to unpack first. The concept of a mechanism, so ubiquitous in the life sciences, has received considerable attention over the past two decades. Naturally, a number of views have been discussed in the literature that differ from each other in various ways,[36] including in the criteria delimiting purported mechanisms. For instance, according to a highly influential account, "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 et al. 2000, p. 3). However, for our purposes and in line with Love and Nathan's stated intention, we may refer to the core conception[37] of a mechanism, according to which "a mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon" (Illari and Williamson 2012, p. 120).

Similar treatment may be applied to the related notion of mechanistic explanation. While there are now many diverse accounts of mechanistic explanation,[38] Love and Nathan remain uncommitted to

[35] Similarly, in Frigg and Hartmann (2020), Frigg presents the usual distinction, but in later work he also develops a more detailed account (Frigg in progress).

[36] See, e.g., Nicholson (2012), Levy (2013) and Andersen (2014a, 2014b) for excellent analysis and overview.

[37] Also known as the minimal conception, see Glennan (2017, p. 17).

[38] Kaplan introduces the model-to-mechanism mapping account (also abbreviated as the 3M account), according to which "a model of a target phenomenon explains that phenomenon to the extent that (a) the variables in the model correspond to identifiable components, activities, and organizational features of the

any particular account in order for their analysis to be as broad as possible. Instead of insisting on the technical details of this or that account, they focus on the core assumption and "treat mechanistic explanation as the claim that many areas of science explain by decomposing systems into their constituent parts, localizing their characteristic activities, and articulating how they are organized to produce a particular effect" (Love and Nathan 2015, p. 762).

Central to their argument is the claim that

> "If the actual difference-making causes are idealized, they do not show how the mechanism actually works. The dilemma should now be apparent. A practice widely used in describing mechanisms—the deliberate misrepresentation of the productive continuity between difference makers—conflicts with the explicit goal of accurately representing causal relations, which is often taken as the hallmark of mechanistic explanation. The idealization of causal relations demonstrates that these models do not depict how the mechanism actually works. If actual difference makers are represented in such a way that they are not difference makers, according to what is already known about the mechanism, mechanistic explanations appear to fail according to their own criteria" (Love and Nathan 2015, p. 768).

Elsewhere they claim that

> "The widespread use of irreducible abstractions challenges the ideal of descriptive completeness, but it is compatible with the goal of describing how mechanisms actually work; abstractions make the model more perspicuous. Idealizations, in contrast, provide a further layer of complexity as they overtly violate the actuality requirement. The introduction of deliberate misrepresentations in a model clashes directly with the claim that mechanistic representations should represent how systems (or their subcomponents) actually work" (Love and Nathan 2015, p. 770).

And finally,

> "accounts of mechanistic explanation face a problem in accommodating the deliberate misrepresentation of causal relations among components and activities that play a difference-making role in producing the explanandum" (Love and Nathan 2015, p. 770).

In their view – adopted in a brief form from Godfrey-Smith's (2009) view – abstraction concerns "the intentional omission of detail"; furthermore, "abstraction must be distinguished from idealization, the deliberate misrepresentation of detail in a model" (Love and Nathan 2015, p. 763). Thus, Love and Nathan's paper belongs to the set of papers that makes use of the distinction.

Many mechanisms in molecular and cellular biology are ordinarily depicted by means of diagrammatic representation accompanied by a description. These often involve depicting entities by means of

---

target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these (perhaps mathematical) variables in the model correspond to causal relations among the components of the target mechanism," to which he further adds that the "3M aligns with the highly plausible assumption that the more accurate and detailed the model is for a target system or phenomenon the better it explains that phenomenon" (Kaplan 2011, p. 347). This particular account has been challenged by, for example, Chirimuuta (2014), who argues against the presumption that the more details the model provides the better it explains the target phenomenon (see also Batterman 2002 for an earlier argument in the same direction, albeit in a somewhat different context; see also Batterman and Rice 2014; Deulofeu et al. 2019). However, it is not clear that Kaplan may be interpreted as subscribing to such a strong statement since elsewhere he states that abstractions and idealizations are a necessary part of scientific work and that they do not jeopardize the explanatory project (Kaplan 2011, p. 348; see also Kaplan and Craver 2011, pp. 609–610).

various geometrical shapes that are connected to each other by arrows standing for causal relations, accompanied by symbols such as '+' or '-' which suggest that a given causal relation is an activating or inhibiting process, respectively. All these shapes and symbols are intended to stand for certain entities and causal processes which purportedly occur inside (or outside) the cell. With respect to these geometrical shapes, any arrows and pluses or minuses that disregard the particular details and the ways in which the entities and causal relations are instantiated are to be considered an intentional omission of details, as abstractions. Other aspects of models in molecular biology are to be taken as instances of idealization, according to Love and Nathan.

The authors present and discuss three assumptions employed in diagrammatic models of gene expression, all of which purportedly introduce important and intentional misrepresentations into the models. Remember that that these misrepresentations should be understood as instances of idealization and since they represent the difference-making factors in a distorted fashion, they cannot provide mechanistic explanations.

Let us look in turn at the modeling assumptions discussed by Love and Nathan.

1) Figure 2.1 depicts the process of gene expression which is initiated by a transcription factor binding to its binding site. However, as illustrated by Figure 2.2, the overall picture is actually much more complex. Instead of a single molecule, a transcription factor, attaching to its binding site, it is actually a functional molecular complex, which consists of various molecules, which does the binding.
2) Both Figure 2.1 and Figure 2.2 "depict gene expression as triggered by a single transcription factor, or, more accurately, a single complex of molecules—call this functional unit p1. While p1 unquestionably plays a role in the process, it is not a difference maker by itself; its presence (or absence) makes virtually no difference to the outcome. This is because even if p1 was not there, another molecular complex of the same type (p2, p3, . . . , p546, . . .) would take its place" (Love and Nathan 2015, pp. 766–767). Furthermore, the functional units continuously bind and detach. Hence, they do not represent an individual binding event but rather a sequence of events that takes place in time and constantly changes.
3) It is the concentration of a transcription factor in the system that fundamentally contributes to its difference-making role.[39] Thus, the diagrams fail to represent aspects of the system that make the actual difference to the occurrence of the modeled phenomenon. More specifically, the key information is the concentration of a transcription factor relative to a repressor.[40]

[39] The causal role of concentrations is discussed in detail in Nathan (2014).

[40] A repressor is any molecule that binds DNA, resulting in either blocking the binding of RNA polymerase to its promoter region, or blocking its function, which, in effect, blocks the DNA transcription.

Figure 2.1. A simple schema of gene expression. Figure drawn by Filip Měšťánek, after Saurabh Sinha.

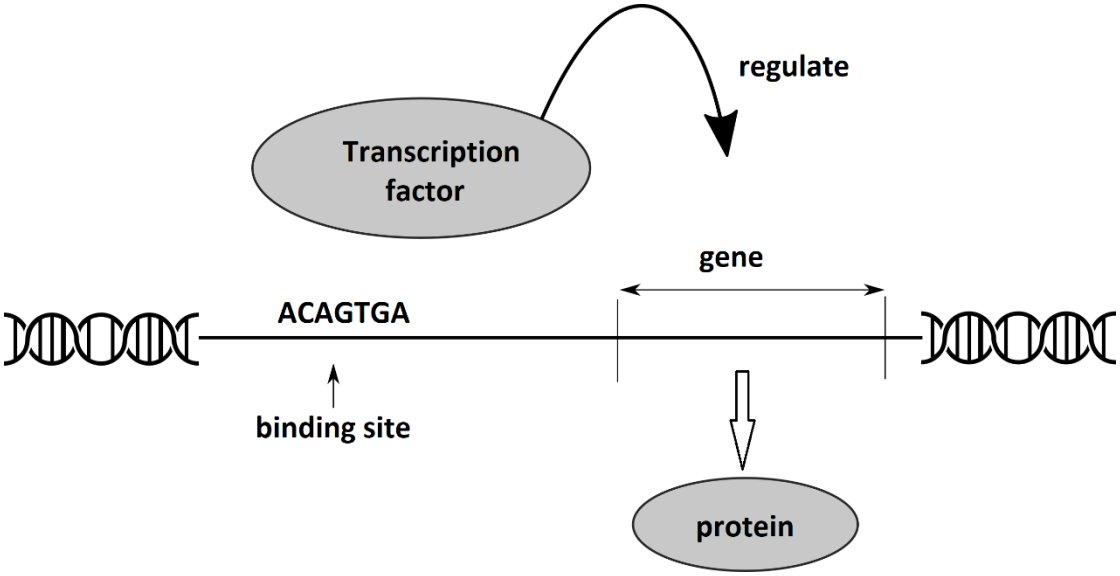
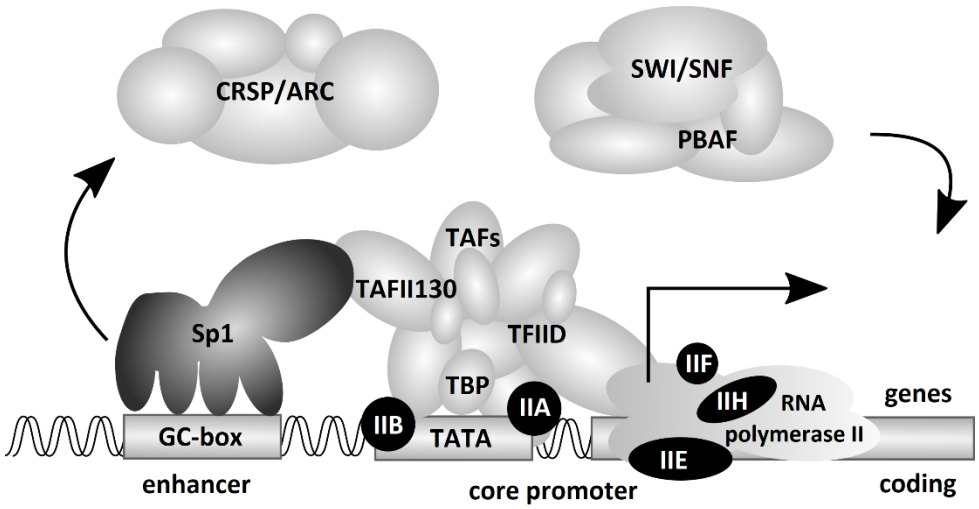


Figure 2.2. A schema of gene expression which depicts the role of many different molecular components in the initiation of transcription. Figure drawn by Filip Měšťánek, after Levine and Tjian (2003).



Love and Nathan are very explicit in stressing that each of the assumptions at hand is an idealization, a misrepresentation of an aspect of the system, and not an abstraction.

2.4. How not to criticize the mechanistic account of explanation in molecular biology

In what follows it will be argued that Love and Nathan's argument suffers from several issues. It paints a confusing picture because the authors are inconsistent with respect to what they argue against as well as to how they use the key concepts; and even if these problems are somehow presumed to be resolved, it still leaves the authors 'empty handed' as it does not provide much support to their intended conclusion. Consequently, their objection to the new mechanistic account of explanation does not stand up to scrutiny, hence, the new mechanists need not worry about this particular objection.

There are two reasons – addressed in turn in Sections 2.4.1. and 2.4.2., respectively – for making the claim that Love and Nathan's analysis paints a confusing picture, and each, even if resolved, fails to provide support to the effect that one would be justified in thinking that the mechanistic account of explanation is in trouble.

2.4.1. Against confusing the distinction

The first problem concerns the fact that despite their insistence on treating all the assumptions discussed in Section 2.3 as misrepresentations (i.e. idealizations), it turns out that they are inconsistent, for they treat one and the same assumption both as an abstraction and an idealization. With respect to the assumption (1) – the binding site binding a single molecule (Figure 2.1) rather than a complex of molecules (Figure 2.2) – Love and Nathan claim that "under normal circumstances, individual molecules do not act as difference makers, but complex functional units do. Thus, the diagram does not "merely leave things out" (abstraction) but "fictionalizes in the service of simplification" (idealization)" (Love and Nathan 2015, p. 766). Clearly, then, we are to think of the concept of a transcription factor as an idealization. However, one page earlier, with respect to the same assumption and the same concept of transcription factor, the authors claim that the "more specific description [i.e., Figure 2.2] of the apparatus for the regulation of eukaryotic gene expression exposes a variety of abstractions that were present in [Figure 2.1]" (Love and Nathan 2015, p. 765). Hence, we are left wondering whether the concept of a transcription factor is an idealization, or an abstraction after all. This is even more pressing when we take at face value the authors' insistence on the claim that "abstraction must be distinguished from idealization" (Love and Nathan 2015, p. 763).

Furthermore, the authors are also inconsistent with respect to their own proposed standards. With respect to the assumption (3) regarding the difference-making role of the concentration of a transcription factor, the authors claim that when "known difference makers are intentionally omitted from the representation" (Love and Nathan 2015, p. 767) we are to understand it as an act of idealization. This, however, is in direct contradiction with their previous definition of abstraction as "the intentional omission of detail" (Love and Nathan 2015, p. 763). Consequently, these contradictory suggestions make the conceptual analysis unclear.

Let us assume, for the sake of argument, that the preceding problems are somehow resolved so that no inconsistencies arise. We may now consider the question of the extent to which the authors are justified in claiming that the assumptions are indeed idealizations. Love and Nathan define idealization as "the deliberate misrepresentation of detail in a model" (Love and Nathan 2015, p. 763) which they adopt from Godfrey-Smith (2009). However, there is no immediate way in which to assess whether or not all the assumptions would in fact count as idealizations in Godfrey-Smith's view. One reason for that is that Godfrey-Smith developed his account in the context of certain kinds of models in evolutionary biology, which, arguably, are very different from the kinds of models used in molecular

biology. Indeed, one should be careful when applying a distinction developed originally in the context of a particular field to another field. Distinguishing abstraction from idealization, as it turns out, is also extremely problematic when the notions are applied to scientific practice. Both these points become clear as soon as one considers the vast literature devoted to making the conceptual distinction between abstraction and idealization, rather than relying merely on the first approximation.

To illustrate, consider again the assumption (3) concerning the omission of concentrations from the model of gene transcription. How should such omission be interpreted? It depends. Discussing an example taken from physics, Martin Jones has the following to say:

> "If a model of a particular fluid flow represents the flow as irrotational when it is not, we can in one sense correctly say that the model omits the rotation involved in the flow. However, such a model omits a certain feature of the real system in a way which involves misrepresenting how things stand in that respect; on the proposal I am putting forward, however, abstractions involve omission *without* misrepresentation. Omission in this restricted sense is, so to speak, a matter of complete silence" (M. R. Jones 2005, p. 175).

According to Jones' view, we may be inclined to construe the assumption as one about which the model remains completely silent (i.e., an abstraction) rather than misrepresenting it. Similarly, Portides (2018) argues that in the process of model building, scientists abstract away from the complexities of the studied system by extracting certain features from it (omission-as-extraction in Portides' terminology) that will serve as the focus of subsequent investigation. Those features which have been retained in a model may then be further modified in an important way, i.e., idealized. The omission of concentrations seems to fit naturally with the notion of abstraction, for this feature is completely missing from the model rather than being modified. Indeed, it is not the case, in contrast with many other cases of idealization, that the parameter of concentrations would be set to zero in this particular model: the model simply does not mention it at all. These remarks are further supported by Love and Nathan who admit that the "explicit descriptions associated with the diagrams (…) do not invoke concentrations" (Love and Nathan 2015, pp. 767–768). Thus, they unwittingly build a case for the assumption to count as an abstraction.

In contrast, other philosophers such as Potochnik (2017) suggest that an omission, sometimes being a failure to explicitly reference a feature, may nonetheless count as an implicit idealization. According to her view, idealization concerns the representation of a target system as if it has a feature it does not, whereas abstraction consists of a straightforward omission which has no consequence for the representation.[41] She then claims that

> "It is important to distinguish this sense of omission—ignoring without representational consequences—from omission in the sense of failing to explicitly reference. It is the former sense that is definitive of abstractions; many idealizations are also omissions in the latter sense" (Potochnik 2017, p. 55).

In support of the proposed distinction, Potochnik provides the example of evolutionary game theory models that hardly ever explicitly state the assumption of the population size being infinite. This assumption allows the modeler to disregard the role of genetic drift. Thus, Potochnik argues, an apparent abstraction is sometimes discovered to be a covert idealization. Understood this way, we may side with Love and Nathan, since their claim that the "explicit descriptions associated with the

[41] Of course, one may ask for clarification of what it precisely means to be of no consequence to representation. Since the validity of Potochnik's views are of little concern to us here, we may simply refer the reader to her original text.

diagrams (…) do not invoke concentrations" (Love and Nathan 2015, pp. 767–768) is in line with treating the omission of concentrations as an instance of covert idealization. Thus, much can be said about both attitudes towards the assumption of omitting concentrations. However, until this question is adequately settled, it may be unfair to proclaim that the mechanistic account should be rejected.

What this shows is that there may be no general agreement among the philosophers writing on the topic of abstraction and idealization on whether a particular assumption should best be viewed as an instance of abstraction or idealization. This can, in part, be due to different epistemic commitments exhibited by different modeling practices in different disciplines such as physics, evolutionary biology and molecular biology.

Consider further the assumption (2): gene expression depicted as being triggered by a continuous event of a single transcription factor binding to the appropriate site as opposed to many transcription factors of the same type and the event being discontinuous. According to Love and Nathan, we should interpret it as an idealization. However, some authors writing on the distinction between idealization and abstraction call for more caution when reaching such conclusions. Describing what is effectively the same type of example – the workings of ribosomes as they are usually depicted – Arnon Levy writes:

> "The ribosome in fact moves back and forth along the mRNA, attaches and detaches, constantly changing conformation. However, it is hard to tell whether standard depictions of this process are idealized or abstract (or both). Do they portray the ribosome as having a sequential, deterministic character, *contra* the realities of ribosomal action? Or are these abstractions that highlight certain functional states and activities of the ribosome while staying silent about others?" (Levy 2018).

Thus, the assumption (2), too, presents a dilemma even for authors who specifically focus on developing a nuanced view of abstraction and idealization.

Moreover, speaking of arrows as means of depicting causal relations in diagrammatic models in molecular biology, Love and Nathan claim that the "arrows simply stand in for causal relations, regardless of how they are instantiated" and that the "typical representation of biochemical components as distinguishable geometrical shapes and the exclusion of known components involves *abstraction*: the intentional omission of detail" (Love and Nathan 2015, p. 763).[42] Causal arrows leave out the particular details regarding how the causal relations are instantiated. In some cases, however, this omission has important consequences for understanding the extent to which a reaction occurs, if at all. Consider a simple model of enzymatic regulation via negative feedback, in which a substrate is turned into a product which feeds back into the pathway and inhibits, thus regulates, the given pathway. Generally, two types of inhibition are recognized: competitive and non-competitive. Holding the concentrations and other conditions fixed, the rates of reaction differ between the two types of inhibitions. Therefore, depending on the particular research question, omitting such details results in withholding potentially crucial information regarding a difference-making factor. Yet, I concur with Love and Nathan that it feels natural to think of causal arrows which stand for any number of causal relations as abstractions. Perhaps the notion of vertical abstraction coined by Mäki (1992), which,

[42] It is rather illustrative of the conceptual difficulties that the authors speak of the geometrical shapes by which the entities are typically represented as abstractions. Whether or a not a particular reaction takes place is influenced by a variety of factors, including – importantly – the particular shapes of the reactants. Thus, although shapes are in fact key difference-making factors for a reaction to occur, they are clearly misrepresented by the diagrammatic sketches. Thus, shapes could potentially be re-interpreted as idealizations.

arguably, also happens to capture much of Levy and Bechtel's (2013) discussion would help to clarify the matters.[43]

Let us summarize the implications of the preceding paragraphs to Love and Nathan's approach. The authors paint a confusing picture by being inconsistent in their use of abstraction and idealization. Arguably, philosophers who develop more nuanced accounts of these notions would give conflicting recommendations as to whether the particular assumptions should be interpreted one way or the other. Additional clues suggest that the assumptions could be understood as certain kinds of abstractions after all. Therefore, the arguments presented by Love and Nathan – hence the conclusion, too – should be taken with a grain of salt. However, another argument in favor of Love and Nathan's approach must be considered before rejecting their analysis completely.

2.4.2. Against confusing what the core objection is

The second problem causing confusion concerns the core objection raised against the mechanistic account of explanation. Recall that *the* problem identified for the mechanistic account pertains to the purported practice of idealizing difference-making factors. In the words of Love and Nathan (2015, p. 768), "if the actual difference-making causes are idealized, they do not show how the mechanism actually works." Although passages like this can be found throughout the text (as evidenced in Section 2.3.), on one place they make the following statement: "we would not expect features that play a central explanatory role to be abstracted away or distorted in a mechanistic description. Yet, in molecular biology, the causal relations responsible for the explanandum are deliberately misrepresented on a regular basis" (Love and Nathan 2015, p. 764). What are we to make of such remarks? On the one hand, we are told that the causal relations in molecular biology are deliberately misrepresented, i.e., idealized, which is why the mechanistic account purportedly fails. That is also why Love and Nathan spill a lot of ink on arguing that the assumptions employed in models in molecular biology are idealizations rather than abstractions (but see the discussion in previous section). On the other hand, though they neither show it, nor argue for it, they seem to casually suggest that abstractions, too, may present a problem for the mechanists. Thus, although they explicitly favor the problem of idealization as a reason for rejecting the mechanistic account, it is not clear whether the problem of abstraction would also be sufficient for them to reject the account.

Let us assume, for the sake of argument, that this confusion, too, is somehow resolved and let us consider what happens if, contrary to Love and Nathan's preferred analysis, we treat the assumptions as abstractions, and if we take at face value their proposal that abstractions present a good enough reason for rejecting the mechanistic account. How would this fare with respect to their own views, and in comparison with other existing views?

Recall that Love and Nathan take issue with what they identify to be the hallmark of mechanistic explanation, namely the goal of accurately representing causal relations (Love and Nathan 2015, p. 768). If causal relations are represented in an idealized – non-accurate – way, then that goal may not be achieved. But the authors argue that abstraction concerns "the intentional omission of detail" and

[43] Love and Nathan are very well aware of the fact that the appropriateness of the chosen level of description must be evaluated with respect to the particular issue at hand (research question, educational purpose etc.). They propose to address this using Weisberg's (2013) multiple model idealization approach. Here we may suggest that the notion of vertical abstraction might serve the purpose better. However, it should also be noted that this particular problem could potentially be re-interpreted as an instance of generalization rather than abstraction, a distinction discussed by Levy (2018).

it "must be distinguished from idealization, the deliberate misrepresentation of detail in a model" (Love and Nathan 2015, p. 763). They say nothing to the effect that abstraction would interfere with the goal of accurate representation, i.e., how mechanisms actually work. Instead, they say precisely the opposite, namely that "the widespread use of irreducible abstractions challenges the ideal of descriptive completeness, but it is compatible with the goal of describing how mechanisms actually work; abstractions make the model more perspicuous" (Love and Nathan 2015, p. 770). However, since they explicitly refer to Godfrey-Smith's (2009) analysis of the notions, one may legitimately ask whether Godfrey-Smith has anything to say on the matter. And he has the following to say: "an abstract description of a system leaves a lot out. But it is not intended to say things that are literally false" (Godfrey-Smith 2009, p. 48). Indeed, as shown in Section 2.2., this view is shared by many of those who use the conceptual distinction between abstraction and idealization as well as those who develop accounts of the distinction.[44] Furthermore, although mechanists have identified research projects in which abstraction proceeds differently in different explanatory contexts – making a representation less abstract by progressively filling in more concrete details (Machamer et al. 2000), or overlooking more detailed descriptions in order to develop more abstract representations such as network models (Levy and Bechtel 2013) –, abstraction is generally construed as harmless with respect to the goal of accurate representation on these accounts.

Thus, even if all the three assumptions are interpreted as abstractions, and even if Love and Nathan do in fact propose that abstractions should somehow, too, pose a problem for the mechanistic account, nothing they say provides support for their conclusion. Instead, most of the clues point in the other direction.

2.4.3. Why should we care?

One key point is that making the precise distinction would not be required if not much depended on it. That is not the case for Love and Nathan, however, who attempt to use the distinction to challenge the mechanistic framework. You will recall that central to their argument is the claim that "the intentional misrepresentation of causal relations, which are the source of explanatory power in a description of a mechanism's components and activities, generates a significant—albeit neglected—problem for the mechanistic framework" (Love and Nathan 2015, p. 762). More specifically, "if the actual difference-making causes are idealized, they do not show how the mechanism actually works" (Love and Nathan 2015, p. 768). However, as has been argued herein, the failure to adequately characterize abstraction and idealization invites trouble for the whole objection to the mechanistic account of explanation precisely because those notions are at the heart of the objection and are presupposed to be clearly defined when they are not. Indeed, unless a stronger foundation is provided, the objection has little traction.

That said, we observe a number of authors expressing sympathies toward Love and Nathan's analysis. Thus, it is worth noting that the detailed analysis of Love and Nathan's approach provided herein may prove illuminating for those engaged in the debate on mechanisms who either explicitly embrace Love and Nathan's views, or follow similar lines of argument.

[44] Note that in the previous section I briefly introduced the example of the negative feedback mechanism and I argued that the arrows representing a causal process may best be viewed as an instance of vertical abstraction. Although such an abstraction is found wanting in context in which a more detailed description is required in order to answer a specific research question, it nevertheless does not say things that are literally false; hence, it does not contradict the received view about abstraction.

For example, Collin Rice references Love and Nathan's work to support his claim that "using idealizations that distort difference-making factors is pervasive in biological modelling—even within mechanistic modelling" (Rice 2019, p. 195). However, I believe I have demonstrated that we have reasons to think that this is something Love and Nathan have failed to show. Similarly, in discussing models of cancer, Anya Plutynski, referring to Love and Nathan, claims that "it is permissible in model building to deliberately represent the system falsely" (Plutynski 2017, p. 131). This is something Love and Nathan failed to prove, so if what Plutynski says is true, it is so for entirely independent reasons. Others (e.g., Halina 2018; van Eck and Mennes 2016) also seem to embrace the core message of Love and Nathan in various contexts.[45] Arguably, this introduces a dangerous precedent which might escalate into a cycle of long-lasting debates, all built on sand.

2.5. Chapter summary

The vast literature on scientific modeling often invokes the concepts of abstraction and idealization. Roughly, two sets of papers may be distinguished: one that develops detailed accounts of the notions, the other that applies these concepts while often referring to the first set of papers. Alan Love and Marco Nathan have relied on these notions when arguing that the mechanistic account of explanation is deeply flawed as it fails to account for the common practice of idealizing difference-making factors.

In this chapter I scrutinized the arguments and examples provided by Love and Nathan and I presented reasons for thinking that their analysis fails to provide support to their conclusion. In particular, I argued that their analysis paints a confusing picture because it is interwoven with inconsistencies regarding (i) how they treat one and the assumption, and (ii) how they apply their preferred definitions.

Setting aside these inconsistencies by assuming none arise, I showed that the assumptions discussed by Love and Nathan are far from clear examples of idealization. Instead, they may very well be interpreted as abstractions. I further showed that philosophers who develop more nuanced accounts would give conflicting recommendations when applied to the particular assumptions discussed herein. In addition to the fact that philosophers disagree on a number of things regarding how precisely to characterize abstraction and idealization, they also develop these notions in a particular disciplinary context. Extrapolating the distinction from one disciplinary context and applying it in another proves challenging, for the range of practices in various disciplines differ considerably.

Furthermore, I argued that Love and Nathan are unclear with respect to the core objection. While they have a beef with idealization not being accounted for in mechanistic explanation, they also seem to suggest that abstraction, too, may be problematic – something for which they do not present any arguments, however. I considered the implications of taking their suggestion at face value and I showed that it does not support their claim that the mechanistic account of explanation is flawed.

Taken together, the analysis provided by Love and Nathan does not provide a sufficient reason for rejecting the mechanistic account. The arguments presented herein should also serve as a cautionary

[45] More precisely, van Eck and Mennes focus on the part where Love and Nathan discuss the use of the multiple-model approach. Halina concerns herself with the representational ideal of completeness. However, the specific details are of little concern to us; the issue at hand is only that these authors touch upon Love and Nathan approvingly without realizing the arguably more fundamental problems discussed herein.

note to those who have embraced the objection to the mechanists, not realizing the fundamental issues underlying such criticism.[46]

It is worth noting that, notwithstanding the intrinsic difficulties of the accounts of abstraction and idealization, these notions may legitimately be invoked in those contexts that do not require a carefully argued analysis. The problem arises only in situations in which these concepts make an appearance in a philosophical debate that calls for careful analysis, such as that on mechanistic explanation. In other words, the concepts of abstraction and idealization concern an important philosophical dispute that should not be treated lightly if we wish to clarify many of the debates which invoke these notions.

[46] However, one should also be wary of interpreting this chapter as a general defense of the mechanistic framework. Indeed, the rich philosophical literature on the mechanistic explanation has many interesting points to offer regarding the tenability of the framework in molecular biology (see, e.g., Skillings 2015). This chapter only meant to show that whatever the means of challenging the mechanistic account of explanation, arguments such as those found in Love and Nathan's analysis are not a good way of accomplishing that goal.

3. Mouse models of cancer: On similarity and representation

3.1. Introduction

Much has been already written on many aspects of research which employs model organisms in order to investigate biological phenomena. Predominantly, the philosophical scholarship has focused on the criteria that guide the choice of model organism, and on the justificatory efforts concerning extrapolative inferences from a model organism to a target system.

Similarly, the concept of similarity has attracted significant attention in the context of the debate on scientific representation, with some attention given to discussing the more specific sense of similarity found in a particular disciplinary context (see, e.g., Sterrett 2017 on the concept of a physically similar system). Focusing on cancer research, and particularly upon cancer immunology, the analysis provided in this chapter contributes to the study of both modeling and similarity by mapping the practices which make use of mouse models. In Chapter 1 I discussed mouse models as systems for generating experimental results that are later used in the construction of conceptual mechanistic models of biological phenomena such as the formation of metastasis. In this chapter, I will provide more details about the specifics of the use of mouse models, thus further elaborate on some of the practices that were presented so far only in a coarse-grained manner.

More specifically, I will be concerned with various kinds of mouse models such as the immunocompetent and immunodeficient transplantable models, genetically engineered models and humanized models. Providing the rudimentary understanding of what is going on in such research, I will then distinguish three research modes: *model selection*, *model creation*, and *model extrapolation*.

The selection of a mouse model is guided by the particular research question at hand, the similarity considerations, and a host of pragmatic and other factors. In model extrapolation, similarity considerations in one way or another are used to justify the extrapolative inferences of the pre-established features of the models. In this sense, much like in model selection, it will be argued that the similarities play a *passive* role. In contrast, model creation amounts to *actively* introducing changes so that a model is made to be similar to a certain degree and in certain respects to the studied phenomenon.

In general, while much has been written on the topic of model selection and model extrapolation, relatively little has been said about creating new animal models. Although the research modes are often intertwined in practice, they are both conceptually and temporally distinct, and as will be argued in some detail, the concept of similarity plays different epistemic roles in each of the modes.

Clarifying these different roles will prove crucial in an argument concerning scientific representation. Most generally, scientific representation has been characterized in terms of one thing standing for another. Thus, a scientific model is a representation of its target system because the model stands for its target. The question, then, concerns the nature of the standing-for relation. What makes a model stand for its target? A number of different accounts have been proposed: structuralist accounts (e.g., French 2003); the DEKI account (Frigg and Nguyen 2020); a variety of inferentialist and pragmatist accounts (Bolinska 2013; Contessa 2007; Knuuttila 2011; Suárez 2004); and the similarity account (Giere 2004; Godfrey-Smith 2006; Mäki 2005; Weisberg 2013), according to which scientists use models to represent their targets by utilizing similarities in certain respects and to certain degrees between a model and its target. Regarding the similarity account, exploiting the relevant similarities is what enables us to learn about the phenomenon of interest by studying its model instead. Despite its

popularity in certain quarters, a wide range of objections have been leveled against the account. According to the objection addressed herein, one must distinguish between the concepts of representation and accurate representation, the latter – but not the former – possibly being grounded in the notion of similarity.

The analysis provided in this chapter will show, however, that the objection holds only to the extent that one is limited to discussing the evaluative aspect of modeling – model extrapolation. In contrast, model selection and model creation illustrate that similarity judgments play a key role in both establishing and maintaining the representational relation between the model and its target phenomenon.

Section 3.2. shall provide the scientific details, illustrating the complex nature of the cancer research (and particularly the cancer immunology research) in which mouse models, owing to the different ways in which they have been (and are still being) developed, exhibit both numerous advantages and limitations with respect to the particular research tasks and questions. Section 3.3. draws philosophical conclusions from the mapping of the field: it presents the three research modes and discusses the role of similarity considerations in each of the modes. Section 3.4. then applies this analysis to the debate on scientific representation. Section 3.5. provides a summary.

3.2. Mouse models in cancer immunology: Transplantable, genetically engineered, and humanized

Much has already been written about various aspects of the research that is distinguished by its use of model organisms.[47] Some have gone to great length to distinguish model organisms from experimental organisms, the latter being a broader category and much less constrained by factors such as the institutionalization and standardization of the research characteristic of the former (Ankeny and Leonelli 2011, 2020). In what follows, the notion of mouse models will be used to refer to mice that are used to study disease mechanisms and for testing drugs.

3.2.1. The field of cancer biology and cancer immunology

At the turn of the twentieth century, oncology opted for a cytotoxic approach to treating malignancies, and until very recently the standard therapies available to patients consisted of radiation therapy, chemotherapy, and targeted therapies. The idea that the immune system could play a role in providing protection against tumors was long treated with suspicion by the scientific and medical communities. However, promising evidence had begun to be amassed and when the FDA approved the use of cancer immunotherapies, such as Sipuleucel-T for the treatment of prostate cancer in 2010 and ipilimumab for the treatment of melanoma in 2011, the field of cancer immunotherapy finally emerged and, to some extent, revolutionized cancer treatment (Farkona et al. 2016; Mellman et al. 2011).[48] These therapies include cancer vaccines, adoptive cell transfer therapy, CAR T-cell therapy, oncolytic viruses,

[47] The question of whether animal models are models *proper* has recently been discussed. Levy and Currie (2015) have argued that animal models should be regarded as distinct from theoretical models, for the two exhibit different epistemic characteristics. Although Parkkinen (2017) concurs, he takes issue with the specific argument by which Levy and Currie reach their conclusion. For the purposes of this chapter the question can be put aside.

[48] Recall from Section 1.3.1. the fact, the role of the immune system in cancer is much more complex as it plays a paradoxical dual role: it has an anti-tumoral effect but it also promotes tumorigenesis.

and the checkpoint inhibitors for which a Nobel Prize in Medicine was awarded in 2018. Despite all this, success has not been absolute, as some patients do not benefit from the treatment, their numbers greatly varying depending upon both the cancer type and the specific immunotherapy in question. Thus, the prospects of a combination of multiple immunotherapies or a combination with more traditional therapies, including targeted therapies, are being investigated.

The process of tumorigenesis includes constituents which can be cell autonomous / intrinsic (i.e. the effect of mutations and/or epigenetic changes on the cancer cell phenotype), and extrinsic (i.e. the tumor environment) (Weinberg 2014). While the former is accessible via studying cell cultures, investigation of the latter consists of aspects that require more complex model systems (Frese and Tuveson 2007). Indeed, much of our current knowledge of cancer biology, and of immunotherapies both in development and in clinical practice, has been gleaned from research conducted on cell cultures and laboratory mice (*Mus musculus*), which allow for *in vitro* and *in vivo* exploratory and hypothesis-driven experiments, respectively.

As technology has advanced, so too have the mouse models: there are spontaneous tumor models, chemically induced models, virally induced models, immunocompetent and immunodeficient models, genetically engineered models, and humanized models. Different models have their own advantages and disadvantages which have a major impact both upon our understanding of the basic mechanisms and on our evaluations of therapies. The models are all said to be standardized, i.e., they consist of inbred strains of mice,[49] and there are protocols about how to manipulate them. The next sections describe only some of the main types of model and the ways in which they have been manipulated, with most attention given to humanized models as these will figure most prominently in some of the arguments to follow.

3.2.2. Transplantable mouse models

Although certain strains of mice are naturally more prone to developing cancer spontaneously, they often develop only a subset of tumor types and grades (Frese and Tuveson 2007). Research in cancer biology has thus found more use for transplantable models which come in two types, to be discussed in turn: (i) immunocompetent models, and (ii) immunodeficient models.

With the advent of syngeneic mice in the first half of the twentieth century, immunocompetent mouse models could be transplanted with tissues including tumors from other histocompatible mice, i.e., mice of the same strain, without such grafts being rejected. Such transplantation studies have led to many important insights including the confirmation of the role of the immune system in tumor surveillance exhibited by the fact that killed tumor cells can act as a vaccine, thus eliciting a potent response against a re-exposure to viable tumor cells of the same type (Budhu et al. 2014; Zitvogel et al. 2016).

Although primary tumor samples can be transplanted, cancer research most commonly relies on injecting mice with cancer cell lines. These are standardized cell lines that originate from tumors derived from a specific background (i.e., the strain), and that, over the course of passaging under *in vitro* conditions, have acquired features that make them well adapted to cell cultures and that set them apart from their ancestors (Weinberg 2014). For instance, the 4T1 breast cancer cell line is derived from the BALB/c strain (Budhu et al. 2014): injecting 4T1 tumor cells into BALB/c mice then leads to an

[49] Inbreeding is thought to provide genetic homogeneity that limits the possibility of the results being confounded. However, as recently pointed out, the current practice of working with mouse strains is not entirely failproof as sub-strains exist and this fact needs to be better acknowledged by researchers (Enríquez 2019).

aggressive growth, giving rise to a detectable tumor mass by day 11, and to a metastasis by day 20. Cancer cell lines have some undisputable advantages. They are easily obtained, maintained, manipulated, and modified for specific research purposes and they lead to highly predictable outcomes. Transplantation studies also differ by route of injection. Orthotopic transplantation refers to injecting tumor cells into the organ where the original cancer developed. While this may be physiologically more relevant, it may not always be feasible to perform such transplantation, for orthotopic transplantation of some tumor types would be too invasive. Therefore, some research is based on ectopic transplantation instead, i.e., tumor cells are injected into a region outside of its original site.

Immunocompetent transplantable models have several advantages: they are cheap, they have a fully functional immune system, they allow for the rapid screening of drugs and for experiments to be conducted in a timely manner, and they lead to extremely predictable tumor growths. However, there are also several disadvantages that make them poorly realistic models in certain respects: the genetic homogeneity of cancer cell lines does not mirror the genetic heterogeneity found in spontaneous tumors, the tumors grow rapidly, lacking the features of multi-step tumorigenesis and the chronic inflammatory environment so characteristic of spontaneous tumors, and they do not recapitulate the tumor microenvironment. Furthermore, owing to their immunocompetence, they cannot be used to directly study human tumors as these would be rejected by the mouse model.

Despite their limitations, immunocompetent mouse models have been among the most commonly used models in cancer immunology and they have contributed to several important discoveries, such as the identification of immune checkpoints and the use of checkpoint inhibitors (Leach et al. 1996), the immunogenic actions of some chemotherapeutic drugs,[50] and the use of combination therapies (Sanmamed et al. 2016; Zitvogel et al. 2016).

Cancer research has also relied upon immunodeficient xenograft mouse models. An early example of such a mouse was the discovery of a nude mouse, i.e., a mouse lacking fur and thymus due to a mutation in the Foxn1 gene. Much like the immunocompetent mouse, cancer research using immunodeficient mice mostly relies on cancer cell lines (but see below).

In contrast to immunocompetent mice, immunodeficient mice can be engrafted with tumors from other mouse strains (allografts) or from other species such as humans (xenografts). Because immunodeficient mice do not reject xenografts, they can be used to study human tumors *in vivo*, which is why most of the current drugs used in oncology, including targeted therapies, have been explored in these mice (Sanmamed et al. 2016). Yet as regards targeted therapies, these mice have served poorly in predicting the outcomes of human clinical trials (Sanmamed et al. 2016). This is because the human tumor cells grow in an environment consisting of mouse stroma cells, which poses problems for translating the results to humans due to improper heterotypic signaling. Another problem concerns the common use of human cancer cell lines, which, much like the mouse cancer cell lines discussed above, do not recapitulate the phenotype of human tumors. Furthermore, the immunomodulatory effects of the drugs cannot be adequately studied in immunodeficient mice. Hence, the scope for testing immunotherapy is severely limited.

[50] For instance, certain chemotherapeutic drugs have more potent effects in transplantable immunocompetent models than in the immunodeficient nude (athymic) mice which lack mature T cells (Zitvogel et al. 2016).

3.2.3. Genetically engineered mouse models (GEMMs)

Acquired knowledge of some of the tenets of oncogenesis and advances in genetic engineering have made it possible to create mouse models with the transgenic expression of oncogenes and inactivated tumor suppressor genes that give rise to spontaneous tumors. Thus, these genetically engineered mouse models (GEMMs) are not reliant upon cancer cell lines. Such mice have not only provided a platform for additional validation of the finding that chemotherapy works in part due to its immunogenic effect, but also for investigating tumor / immune system interactions by removing some of the essential immune-related genes such as those encoding perforin or IFN-γ and its receptor which accelerates tumorigenesis (Zitvogel et al. 2016). They have also served as preclinical systems upon which oncology treatments including immunotherapy have been tested (Sanmamed et al. 2016).

We have seen that different transplantable models provide different advantages while also exhibiting some disadvantages. The same can be said of GEMMs. The fact that these are spontaneous models means that, in contrast with models reliant upon cancer cell lines, they maintain some of the features of tumorigenesis, plus a tumor microenvironment which displays suppressive characteristics, with T cells expressing multiple checkpoint molecules allowing for the testing of checkpoint inhibitors.

However, GEMMs encounter major challenges owing to the specific ways in which these models are created. There is an array of technologies available for creating transgenic mice, impacting *how* the mouse is modified, *where* it is modified, and *when* it is modified.

For instance, using recombinant DNA technology, GEMMs can be created by the direct injection of fertilized oocytes (germ cells), or by the use of the lentiviral transduction of embryonic stem cells. One problem is that both the expression levels and the cell tropism of the transgene may not completely reflect the expression levels and cell tropism of the endogenous gene. In other words, the transgene is often both overexpressed and expressed ectopically. This is because the promoter fragments in the transgene typically contain only the minimal sequence necessary for expression but not all the regulatory sites as in endogenous genes (Frese and Tuveson 2007). To avoid this problem, researchers may use tissue-specific promoters in order to limit the expression to cells of a particular tissue. Under these circumstances, all mammary epithelial cells, for instance, will express the *Erbb2* transgene, which will simultaneously give rise to multiple neoplastic lesions (Zitvogel et al. 2016). Consequently, the immune system is overwhelmed, and the effects of a therapy may be confounded: Zitvogel and colleagues report that the MMTV-*Erbb2*-induced breast cancer exhibits no measurable immune system impact in response to chemotherapy; indeed, a good response to chemotherapy is observed even in RAG2 knock-out models, which conflicts with the clinical observation that favorable responses are associated with tumor infiltrating lymphocytes. Other problems concern the random site of the integration of the transgenes, which may result in chromosomal positional effects. It has also been possible to induce oncogenic mutations in adult mice in only a subset of cells which better mimics the features of human cancers (Sanmamed et al. 2016). As seen in transplantable models, different strains used in the developing GEMMs have different biological predispositions, meaning that the results obtained across the spectrum may differ accordingly (Frese and Tuveson 2007). Finally, GEMMs typically require longer follow up as tumors appear after several months and experiments take many additional months (cf. transplantable models).

3.2.4. Humanized mouse models

As noted above, both transplantable and genetically engineered mouse models have proven useful in both basic and preclinical research. However, many of these models exhibit an environment which is

only poorly representative of the condition in humans. Given that there is a growing need for animal models to serve as systems which could overcome some of the limits of the previously discussed models, and upon which *in vivo* studies of human cells and tissues could be conducted, some researchers are now investing in developing mouse-human chimeras, so-called humanized mouse models. Humanized mice can be defined as "immunodeficient mice engrafted with haematopoietic cells or tissues, or mice that transgenically express human genes" (Shultz et al. 2007, p. 118). Thus, humanized mice are a specific kind of transplantable model, and some such models also make use of the transgenic approach. In addition to studying cancer, including tumorigenesis, metastasis and cancer therapy, humanized mice have also contributed to advances in studies on human immunity, infectious diseases, and regenerative medicine (see Walsh et al. 2017 for an extended overview).

Together with advances in engraftment techniques, three breakthroughs have made the development of humanized mice possible (Shultz et al. 2007). The first concerned the discovery of the autosomal recessive mutation of the *PRKDC* gene (protein kinase, DNA activated, catalytic polypeptide) in the CB17 mice, leading to a complete lack of mature B and T cells which then manifests in the form of severe combined immunodeficiency (SCID). However, the success of human-cell engraftment is limited both by the presence of an innate immune system, including natural killer (NK) cells, and by the spontaneous generation of B and T cells – known as leakiness – during the aging of the mouse. The latter issue is resolved by working with RAG1 and/or RAG2 deficient mice which never develop functional B or T cells. The second breakthrough was the development of immunodeficient non-obese diabetic mice (NOD mice). Soon thereafter researchers realized that the subsequent crossing of the *scid* mutation onto a NOD background (NOD-*scid* mice) led to a significant improvement in the engraftment of human peripheral blood mononuclear cells (PBMCs) compared with strains such as the C57BL/6-*scid* mouse. In part, this is because NOD mice have an impaired innate immunity.[51] Finally, at the turn of the twenty-first century, the third breakthrough came about in the form of the targeted mutations at the interleukin-2 receptor subunit gamma (*IL2RG*) also known as the common gamma chain, which results in the impaired signaling of a number of cytokines and completely prevents NK-cell development. Mice bearing this mutation allow for the greatly improved engraftment of human tissues.

Multiple strains of mice bearing the *IL2RG* mutation have been developed. It is noteworthy that differences in the genetic backgrounds of mouse strains influence the extent to which human tissues are successfully engrafted (Shultz et al. 2012).

There are also at least three ways of engrafting a functioning human immune system into immunodeficient mice bearing the above mutations. Additionally, each approach can proceed via a different route of injection, which often leads to different results. The simplest approach is to engraft into an immunodeficient mouse both human peripheral blood lymphocytes and tumors, ideally from the same donor, thereby creating an *immuno-avatar* (Sanmamed et al. 2016; Zitvogel et al. 2016). Owing to their human immune-tumor interface, immuno-avatars have been used to identify and screen antihuman CTLA-4 monoclonal antibodies and other immunomodulatory drugs that can activate human PBMCs. However, the use of immuno-avatars is constrained by the onset of severe human xenograft versus host disease (xGVHD) just a few weeks after the engraftment. Thus, unless the onset of xGVHD is delayed, such as by depleting CD4 positive cells from the PBMCs beforehand, the time window for experiments is rather limited (Sanmamed et al. 2016).

[51] The deficiency involves mouse macrophages and mouse complement respectively impairing the phagocytosis and lysis of human cells.

Another way to create humanized mice is to engraft human hematopoietic stem (and progenitor) cells (HSPCs). The quality of the engraftment depends upon many factors, including the particular recipient mouse strain (Shultz et al. 2012), or the site from which $CD34^+$ HSPCs have been isolated, whether from cord blood, bone marrow, peripheral blood, or fetal liver (Sanmamed et al. 2016).

Although this technique of humanizing mice represents an important stepping-stone towards addressing some of the limitations of previous models, several additional issues have emerged. To become mature T cells, the progenitors must undergo both negative and positive selection in the thymus. Because the human-derived T cells are trained on mouse thymus cells and later interact with human-derived antigen-presenting cells which express different sets of MHC molecules, the T cells do not perform well. Indeed, their functionality has been found to be limited, with minimal proliferative potential and the tendency to become anergic upon activation. Likewise, human-derived B cells originating in bone marrow fail to complete the maturation process in the spleen, and although they can produce IgM antibodies, they do not perform class switching (Morton et al. 2016). To overcome these problems, researchers have transplanted human fetal bone, fetal liver, and thymus tissue beneath the kidney capsule, thus creating the so-called BLT mouse, followed by the engraftment of HSPCs. As a result, these mice do form functional mature T and B cells. However, practical issues, such as the need to generate human organoids and to obtain enough primary tissue for large cohorts, limit the utility of these mice. Additionally, the key to a fully functioning immune system lies in a host of other factors, including species-specific cytokines, growth factors, and homing molecules (Sanmamed et al. 2016). Although it is possible to administer exogeneous human cytokines, it often leads to non-physiological concentrations, causing unnatural behavior in the immune cells. Mice engrafted with human HSPCs also generate human NK cells but their ability to kill their targets is impaired, in part because mice do no express human MHC (i.e., HLA) molecules. Finally, unless exogenous interleukin-7 and lymphotoxin receptor agonist are administered during crucial developmental steps, these mice also exhibit deficiencies in the development of (peripheral rather than mesenteric) lymph nodes and the structure of lymphoid tissues is poorly organized (Shultz et al. 2012).

A natural move forward is to use transgenic technology, the third and last approach to creating humanized mice to be discussed here. Again, multiple technologies are employed, generating diverse results which also depend upon the particular mouse strain used. There are three main technological approaches to delivering human species-specific factors into the genome of mice to enhance human hematopoiesis and immune system development and function, as well as ways to inactivate mouse genes in a target-specific manner. These approaches include: (i) transgenic expression of cDNA constructs driven by tissue-specific or ubiquitous promoters; (ii) transgenic expression of bacterial artificial chromosomes (BACs); and (iii) knock-in technology (Shultz et al. 2012) which currently includes the use of CRISPR-Cas9 technology.[52]

[52] Shultz and colleagues (2012) also provide more detailed discussion: the problem with (i) is that the method often results in the gene being expressed at non-physiological levels and without temporal control, which adversely affects the development and function of the human immune system in mice. In contrast, the use of (ii) leads to expression at physiological levels in an appropriate functional and developmental context. Note that if the mouse (homologous) genes are not silenced, they too will be expressed. Finally, with the increase in the availability of embryonic stem cells taken from the strains with a NOD background in which mouse genes have been knocked out, option (iii) generates mice in which the mouse gene is replaced by its human counterpart. Generally, the knock-in approach is defined as follows: "The introduction of a transgene into a precise location in the genome, rather than a random integration site. Knocking in uses the same technique of homologous recombination as a knockout strategy, but the targeting vector is designed to allow expression of the introduced transgene under the control of the regulatory elements of the targeted gene" (Shultz et al. 2012, p. 790).

One alternative to the BLT mice discussed above is to construct mouse models that transgenically express a common HLA allele in their thymus and that are engrafted with HSPCs with the matching histocompatibility alleles, thus facilitating the maturing process of T cells by allowing for the selection process to take place as well as securing their functionality in the periphery. However, these efforts have been met with mixed success, further showing the complexity of the biology involved (Morton et al. 2016).

To ensure that a fully functioning human system is in fact established in the immunodeficient mouse host, researchers have also introduced into these mice human transgenes coding for some of the key cytokines such as THPO and GM-CSF, using the various transgenic techniques outlined above. Although this has led to a significant improvement in the development of the cells of myeloid origin in particular, adverse events have also been observed (Shultz et al. 2012). Still, such a procedure has recently been utilized by knocking in human genes coding for M-CSF, GM-CSF, IL-3 and thrombopoietin, thus creating the MITRG mice (Rongvaux et al. 2014). The same group also created the MISTRG mice using the bacterial artificial chromosome (BAC) transgene which codes for human SIRPα which binds the CD47 molecule expressed on human cells, the interaction of which results in suppressing phagocytosis. Along with functional B and T cells, these mice also reconstituted functional NK and myeloid cells.

The last mouse model to be described – the patient-derived xenograft mouse model (PDX) and its humanized version – builds on much of the previous discussion and while the model improves on some of the limits of other models, some challenges remain, and novel ones arise. In contrast with other xenograft models which most commonly rely on working with cancer cell lines, PDX models[53] are immunodeficient mice that become host to freshly resected tumor samples obtained from patients (Decker et al. 2017; Sanmamed et al. 2016). Because the PDX mice are generated by the surgical transplantation of human tumor samples, many of the features of the tumor microenvironment are kept intact, which allows for immunotherapies to be tested (but see below). To increase mouse cohorts, researchers take advantage of serial passaging, i.e., the expansion of tumor-bearing mouse cohorts by transplanting the engrafted human tumor from the P0 generation to the next. However, with the increase in mouse cohorts the original tumor sample is diluted and eventually lost, limiting the capacity to investigate the effects of immunotherapies. The fact that PDX models depend on tumor samples rather than on the easily obtainable and maintained repertoires of cancer cell lines presents an obstacle due to the relative scarcity of tumor material, resulting in fewer mouse cohorts. Furthermore, some tumor samples are difficult – even impossible – to obtain owing to the extremely invasive surgical procedure required for accessing the tumor. Working with tumor samples is, in some respects, also more challenging than working with cell lines. PDX models also suffer from many of the same problems as other common xenograft models, such as the particular transplantation method (ectopic as opposed to orthotopic), and the lack of a fully functional immune system, which puts constraints on the adequate investigation of the tumor-immune system interaction and, therefore, of immunotherapies.

To overcome the latter problem, efforts have been made to create humanized PDX models – also called immune-PDX models, or iPDX for short (Sanmamed et al. 2016), and particularly by transplanting the HSPCs and tumor sample from the same donor in order to limit the confounding resulting from the tumor being rejected due to the tumor-HSPCs histological incompatibility (Morton et al. 2016). However, the iPDX's greatest advantage is also its greatest disadvantage, for the tissue available for

[53] Note that although PDX models are also sometimes referred to as mouse avatars, some researchers (see, e.g., Figure 1 in Sanmamed et al. 2016) reserve the term for other types of mouse models such as the mice humanized by PBMCs engraftment, i.e., the immuno-avatar discussed previously. The non-humanized version of PDX models has recently been the topic of an extended philosophical analysis (Green et al. 2021).

conducting studies is scarce and, in contrast with PDX, iPDX are confined to biopsy samples as serial passaging is not desirable in these settings. A related issue concerns the use of these models in the context of personalized and precision oncology: they should allow for patient-specific drug testing in real-time but since creating models, followed by the research, takes time which the patients may not have, temporal challenges arise (see Green et al. 2021). Although the iPDX mouse models provide an improvement over the PDX in that the former is equipped with the human immune system, the immune system's functionality remains limited unless many other previously discussed modifications are introduced, such as the introduction of human adhesion molecules for more adequate trafficking patterns.

3.2.5. Future challenges

Despite the clear success of the current models, much still remains to be done. Two issues pulling in somewhat different directions can be identified. First, while the models in use are appropriate for addressing many questions of interest, they need to further evolve so as to correct the remaining deficiencies and thereby improve the engraftment and function of the human immune system (Shultz et al. 2012; Zitvogel et al. 2016). For instance, it has been established that the microbiome plays a major immunomodulatory role and influences the effectiveness of immunotherapies such as checkpoint blockade. However, the mouse models harbor their own microbiome, thus possibly confounding the results of preclinical testing. Zitvogel and colleagues suggests that the gut microbiome should be humanized, ideally in a patient-specific way, to improve the iPDX models. The second research direction also calls for an improvement but with a twist: Decker and colleagues note that experiments take place in the sterile environment of labs, meaning that the genetic and environmental variability found in normal populations is lost and the exposure to normal commensals is limited, all of which obfuscates the complexities and the variability of the multifactorial neoplastic disease (Decker et al. 2017). Thus, the way ahead may incorporate large, outbred cohorts into cancer research.

3.3. Introducing similarity: model selection, model extrapolation, model creation

Having provided some exposition of the mouse models in cancer biology and cancer immunology, we now move on to considering the role that similarity considerations play in working with these models. Three research modes will be identified and distinguished from one another: model selection, model extrapolation and model creation. Although these research modes are often more or less intertwined in practice, they can be conceptually disentangled and analyzed. Indeed, arguably the philosophical literature has been focused on discussing one or another. Moreover, whereas much has been written on the topic of model selection and model extrapolation, relatively little has been said about creating new animal models.[54] By focusing separately on each mode we can better flesh out the diverse roles played by similarity considerations.

[54] In fact, some authors have explicitly denied that model organisms are being created. For instance, with respect to model organisms, Weisberg (2013, p. 16, italics added for emphasis) claims that "although they are not *constructed*, like the San Francisco Bay model, they are concrete systems that resemble concrete targets".

3.3.1. Model selection

The selection of mouse models is driven by several considerations: (i) the research question at hand, or what I term 'adequacy-to-research-question',[55] (ii) the pragmatic and situational constraints, and (iii) the relevant similarity considerations.[56] Depending on the question, scientists may use one model over another. For instance, if the aim is to test immunotherapies, researchers will want to rely on mouse models in which the particular immunotherapy can be tested. As we have seen, immunodeficient transplantable models using cancer cell lines would be an inappropriate choice for that particular task.

Model selection is also constrained by many other factors beyond adequacy-to-research-question. Recently, Dietrich *et al.* (2020) have provided an extensive discussion of what characterizes a good or useful organism for a given research interest, identifying over 20 criteria clustered into several categories, often interrelated but also in tension and traded off against one another (see Table 3.1).

Table 3.1. Criteria for organismal choice. Adopted from Dietrich *et al.* (2020).

Cluster	Criteria
(A) Access	(1) Ease of Supply
	(2) Phenomenal Access
	(3) Ethical Considerations
(B) Tractability	(4) Standardization
	(5) Viability and Durability
	(6) Responsiveness
	(7) Availability of Methods and Techniques
	(8) Researcher Risks
(C) Resourcing	(9) Previous use
	(10) Epistemic Resources
	(11) Training Requirements
	(12) Informational Resources
(D) Economies	(13) Institutional Support
	(14) Financial Considerations
	(15) Community Support
	(16) Affective and Cultural Attributes
(E) Promise	(17) Commercial and Other Applications
	(18) Comparative Potential
	(19) Translational Potential
	(20) Novelty

Most, if not all of these criteria, are also applicable to the specific context of mouse models in cancer biology and immunology. Thus, these criteria guide not only the choice of model organism in general (e.g., the choice between zebrafish and rodents), but also the very particular selection of mouse models (e.g., between transplantable and genetically engineered). For example, consider the criterion of *phenomenal access*: transplantable immunodeficient mouse models allow neither for investigation

[55] To some extent, this should be reminiscent of the "identification of targets of modeling" discussed by Huber and Keuck (2013). The concept of adequacy-to-research-question is also akin to Parker's (2020) notion of adequacy-for-purpose as well as to Bolinska's (2016) accuracy-for-a-purpose.

[56] Note that the order in which these considerations will be discussed does not necessarily reflect some rigid order in which scientists actually reach conclusions. Actually, these points are usually interdependent.

of the role of the immune system in surveilling tumors, nor for testing immunomodulatory drugs such as checkpoint inhibitors. Likewise, the fact that engrafting PBMCs brings about the onset of xGVHD within a few weeks, thus limiting the time window for testing checkpoint blockers, showcases other practical and epistemic constraints captured by the *phenomenal access* and *viability and durability* criteria. The *ease of supply* of standard immunocompetent mice allows for running experiments and screening drugs on large enough cohorts, which generally presents more difficulties to various kinds of humanized models. Although humanized mice are said to provide better *translational potential*, they also put pressure on *financial considerations* as they are significantly more expensive, which means that many laboratories cannot afford them, and so *institutional support* is limited. Although the criterion of *previous use*, that is, for example, how to feed the mouse or how it reacts to the laboratory environment, does not present a significant difference between humanized and other mouse models, the *training requirements* and *informational resources* concerning the specificity of these models as opposed to other types of models does present a difference, such as what the limits of particular humanized models are. In the context of cancer mouse models, and particularly cancer immunology mouse models, s*tandardization* concerns not only the inbred mouse strains[57] but also the particular method applied to them: the type of transplantation of tumor (ectopic or orthotopic) or PBMCs/HSPCs (intravenous or otherwise) etc., all of which affect the experimental data. Humanized models also present *novelty* and can be considered as emerging models. Thus, selecting the right model for the given task at hand can be quite challenging.[58]

One important criterion, thus far denied explicit consideration but nevertheless underlying many of the above criteria, is similarity or resemblance judgment. The crucial role of similarity considerations in model selection is widely acknowledged across both the philosophical and scientific communities.[59] Dietrich and colleagues are quite clear when writing that "most commonly, organisms are chosen because of their physiological or genetic resemblance to humans, the presence of similar mechanisms in both species, or due to high rates of incidence of a given disease of interest" (Dietrich et al. 2020, p. 8). The same can be said with respect to choosing a particular mouse model, given the specific research question. In other words, the mouse model is chosen for its presumed similarity in relevant respects to the particular aspect of the cancer being studied. To select a mouse model for investigating one's research project is to use the mouse as a representation of a particular phenomenon (or one of its aspects). Although no intrinsic feature of mouse models dictates which model will be used for representational purposes - because other factors influence model selection (see above) - the intrinsic features do have an important epistemic role in choosing a model with which to work.

As noted above, the intrinsic features of the model that inform the selection process are features that are considered to be relevantly similar to the features of the target, given the research question at

[57] Complicating the matter even further, Enríquez (2019) recently argued that unknown to many, standard strains often form sub-strains which differ from each other by, for example, certain metabolic features.

[58] The philosophical literature has described a number of such cases in detail, including the use of rabbits in the study of atherosclerosis in humans (Parkkinen et al. 2017), the use of rodents in research on alcoholism (Ankeny et al. 2014 arguing that both the organism and its environment - its 'situatedness' - must be taken into account), and the use of an inferior mouse model for human Down syndrome when a complete genetic mouse model is available (Hardesty 2018).

[59] Some historians of science couch their descriptions of historical episodes along the same lines. Consider, for instance, the words of Frederic Holmes, who thus describes the work of Marcello Malpighi, the 17th century Italian physician, and the rise to prominence of frogs as experimental models: although "frogs had simpler lungs than mammals", because they "looked *similar*, were *similarly* placed, and were *similarly* connected to blood vessels and trachea, there was no reason to doubt that their basic structure and functions corresponded to those of the lungs in 'higher' animals" (Holmes 1993, p. 315, italics added for emphasis). See also Ankeny and Leonelli (2011) for discussion of this case.

hand. Consequently, it would be wrong to think that a single model that exhibits relevant similarities for answering a particular question is also suited to answering another question. As Shultz and colleagues put it

> "Any one specific model will not be optimal for addressing the myriad of questions that might be considered, and it is important not only to choose the appropriate model system for the specific question at hand, but also to be innovative in formulating questions and experimental designs to provide valid data that can be properly interpreted using the individual models" (Shultz et al. 2012, p. 787).

The same can be said with respect to choosing a particular humanized mouse model from the available repertoire, given that different engrafting methods provide different settings as shown in Section 3.2. To again put it in the words of Shultz and colleagues: "Depending on the question, the investigator will need to choose the appropriate human immune system-engrafted mouse for their studies" (Shultz et al. 2012, p. 787).

On a related note, it is taken for granted that any model will exhibit countless dissimilarities with respect to the target, but as long as these dissimilarities are considered to be irrelevant to the task at hand, they are deemed of no significant importance.[60] Of course this says nothing about the likely possibility that it may later turn out that seemingly irrelevant traits are in fact relevant. For instance, although it was long assumed that chemotherapies have a purely detrimental effect on the immune system, which was thought to play no role in cancer surveillance, it turned out that some chemotherapies have an additional anti-tumor effect by being immunogenic. Recall that some genetically engineered models overexpress oncogenes in a manner that overwhelms the immune system, with chemotherapy showing no benefit over RAG2 knockout (immunodeficient) mice. This example shows that even the discovery that a model has limitations may not result in abandoning the model. As long as the model still captures at least to some degree some relevant aspects of the phenomenon, researchers may decide to continue using the model, especially if factors such as ease of access can be traded in for similarity considerations. Genetically engineered models can still serve as extremely useful sources of information, even if some specific models may confound some specific results. Moreover, even when there is a is readily available model that is more similar in relevant respects than another model, such as a PDX model's superiority over a simple immunodeficient transplantable model for the study of human cancer, it is not necessarily prioritized over a model that exhibits similarity to a lesser degree, as other considerations are factored in when deciding, such as the particular research question and the pragmatic and other constraints (see table 3.1). So, although the degree of similarity might be traded for other virtues such as ease of supply, similarity considerations do inform the process of model selection in important ways, for if a model turns out to completely miss on any relevant similarity, it will be abandoned no matter what the other benefits may be.

Finally, the practice of considering similarities in the process of model selection can be viewed as *passive* in the sense that the similarities entering the decision-making process of choosing a mouse model are pre-established rather than actively introduced, that is, the similarities in question had been explored and established prior to the point at which a researcher selected the model with which to work.

[60] However, some biomedical research is motivated precisely by studying dissimilarities using negative models, which are interesting because they do not exhibit the disease in question, with finding out *why not* offering possible insights or solutions to the disease found in humans (Dietrich et al. 2020; Green et al. 2018).

3.3.2. Model extrapolation

Arguably, philosophical literature has paid the most attention to the issue of extrapolation, also known as external validity, that is, the process of making inferences from an animal model to the target. The aim of such scholarly work is to either provide an explicit description of the ways in which these inferences are justified in practice, or to suggest criteria for improving the justificatory efforts. The problem of extrapolation or external validity can be summed up as follows: "Evaluating external validity (…) requires evaluating whether the complex of relevant mechanisms in the target population is sufficiently similar to that in the study population" (Parkkinen et al. 2018, p. 17). This concerns both basic and translational research. For example, using findings obtained from pre-clinical studies to project the efficacy and safety of a potential anticancer drug in human trials is known to suffer from high rates of failure: it is estimated that around 85-89% of potential anticancer drugs ultimately fail to gain approval (Sanmamed et al. 2016). More importantly, there are well documented cases in which experiments conducted on animal models showed promise but had tragic consequences in first-in-human trials (Lemoine 2017; see also Parkkinen and Williamson 2020).

While philosophers appear to be in agreement with respect to the general claim that the success of extrapolation depends upon some sort of similarity between the surrogate and its target, they disagree on the particular manner in which the justification proceeds.[61] Indeed, several such accounts have been proposed: enumerative induction, comparative process tracing (Steel 2008), phylogenetic inference (Bolker 2009; Levy and Currie 2015; Weisberg 2013), robustness reasoning (Parkkinen and Williamson 2020),[62] theoretical chimeras (Lemoine 2017), and experimental manipulation (Maugeri and Blasimme 2011; Piotrowska 2013). These different accounts have been developed against the backdrop of different research projects. Thus, it is possible that while a particular account of extrapolation may reasonably well capture what is going on in a particular context, it may fail in another: for instance, the view that model extrapolation generally proceeds via phylogenetic inference has been challenged, using the example of engineering models in biomedical research (Maugeri and Blasimme 2011; Parkkinen 2017).

Although all mouse models discussed here have provided useful insight into cancer biology and generated key pre-clinical data - think of the checkpoint blockers tested in transplantable models - researchers are also well aware of their limits. For instance, Budhu and colleagues write:

> "It is widely accepted that mouse models are able to provide useful pre-clinical and mechanistic information about novel immunotherapies and cancer therapies. However, an argument that is very often brought up is that animal studies are uninformative because they are not predictive of results in humans. Inadequacies in experimental designs may account for some of these failures. As tumor biologists select models to evaluate an immunotherapy or ask a specific question, it is vital to insure that the proposed models recapitulate and mimic the human disease as closely as possible, ensuring that pathology, metastatic potential, stage of disease, extent of tumor burden, hormone responsiveness and immune suppression are adequately and faithfully recapitulated in the animal models corresponding to each studied

[61] However, note that although the "translational potential of experimental organisms can (…) stem from similarities", it can also stem from "differences to human physiology" (Dietrich et al. 2020, p. 8).

[62] See Parkkinen and Williamson (2020) for an assessment of the four approaches mentioned above. See also Baetu (2016) and Ankeny and Leonelli (2020, p. 56) for a criticism of phylogenetic and other traditional approaches.

> cancer type. It is also important to take into account the predictability and limits of each model when translating mouse experimental data into clinic" (Budhu et al. 2014, p. 50).

In part, the relative scarcity of translational successes may be explained by an overreliance on models that recapitulate only poorly the key features of cancer. For example, as discussed in Section 3.2.2., the problem with transplantable immunodeficient mouse models is that the human cancer cell lines, much like their mouse counterparts, do not recapitulate the phenotype of human tumors, and they grow in an environment consisting of mouse stroma cells which poses problems for translating the results to humans due to improper heterotypic signaling. In the words of Morton and colleagues,

> "Both cultured cells and mouse xenografts grow in an environment highly dissimilar to that of their originating tumor, frequently resulting in promising treatments that are ultimately clinically ineffective. The development of highly immunodeficient mouse strains into which human immune systems can be engrafted can help bridge this gap" (Morton et al. 2016, p. 6153).

Thus, it is believed that the development of humanized mouse models which are made to be more similar to the intended target may help cross the translational gap.[63]

There is an apparent connection between model selection and model extrapolation. When selecting a mouse model in the translational context, one often does so by considering the likelihood that the findings will prove relevant to the target. Both model selection and model extrapolation rely on pre-existing features and pre-established similarities. More specifically, in model extrapolation it is the consequences of the pre-established similarities that are the central focus. In this restricted sense, model selection and model extrapolation are 'static' as both draw on pre-established features.[64] Nevertheless, model selection and extrapolation are conceptually and epistemically distinct for at least two reasons. First, model selection takes place prior to conducting research whereas model extrapolation concerns the practice of projecting the obtained results onto the target. Second, the act of selecting a mouse model for studying a given phenomenon differs from the act of justifying the extrapolation from the mouse to the target phenomenon. The fact that scientists consider certain relevant mouse features to be similar to those of the phenomenon does not imply that the purported similarities in fact hold, and even if they do, they may still fail to support the extrapolative inferences, for the similarities may prove to be shallow, or their effect may be masked by other causally relevant aspects present in the mouse and not in humans or *vice versa*. This point can be rendered more explicit by distinguishing between presumed *representational accuracy* concerning the similarities guiding model selection and *predictive accuracy* which pertains to the assessment of the extrapolative inferences, i.e., the extent to which the obtained results in question mimic the features of the target system.[65]

3.3.3. Model creation

Manipulation is at the heart of creating mouse models. Taking mice from the wilderness and into the laboratory environment, as used to be the case, and breeding genetically homogenous cohorts (mouse

[63] Piotrowska (2013) proposes several criteria (heuristics) which, taken together, should guide scientific judgment regarding the extent to which the humanized mice are similar enough to the modeled disease to justify the extrapolative inferences.
[64] See also Ankeny and Leonelli (2020, p. 56), who speak of most philosophical accounts of animal model extrapolation as static, in contrast to what is usually going on in research-related processes.
[65] I am indebted to Sara Green for sharpening my thinking here.

strains) essentially amounted to creating various (syngeneic) mouse models. The development of cell lines and their subsequent use in transplantation studies led to the creation of immunocompetent transplantable models. Similarly, exploiting naturally occurring mutations which gave rise to immunodeficient mice have cleared the way for creating mouse models capable of hosting xenografts. Advances in genetic engineering have enabled the creation of new and specific mouse models. In fact, the creation of specific knock-out mice, for instance, and their comparison with their wild type counterparts, forms part of the contemporary research routine. The process of creating humanized mouse models is a combination of many of these general methods.

In contrast with model selection and model extrapolation, model creation is a process of introducing targeted changes that give rise to new mouse variants rather than working with what is given. Rather than making use of pre-established features, creating models amounts to actively adding new, removing old, or modifying existing features. Thus, model creation may better be characterized as an *active* process rather than something *passive*. Such a processual nature has been emphasized before in the philosophical literature, such as in (Ankeny and Leonelli 2020; Atanasova 2015; Huber and Keuck 2013; Lemoine 2017; Maugeri and Blasimme 2011; Parkkinen 2017; Piotrowska 2013).

As noted above and notwithstanding the differences, there are also connections between model creation and the two other research modes. The process of creating a specific model does not happen in a vacuum; it is guided by the particular research task one is trying to address. Thus, model creation, much like model selection, ought to be characterized in terms of adequacy-to-research-question. Indeed, as Shultz and colleagues write, "experiments using humanized mice, or any animal model system, need to be designed to address a mechanistic question rather than attempting to fully recapitulate the human biological process or pathology" (Shultz et al. 2012, p. 796). Among other things, Section 3.2. demonstrated that there are "different technological approaches for the engraftment of a functional human immune system in these immunodeficient mouse models, each with distinct advantages and caveats" (Shultz et al. 2012, p. 787). Recall, for instance, that one option is to engraft PBMCs, which results in the early onset of xGVHD. If the research task requires a longer follow up, then human-PBMCs-bearing mice would not be the ideal model to create, given the alternatives.

It should be noted that, in practice, model creation is often clearly connected with model extrapolation (see, e.g., Atanasova 2015), since often one of the reasons for introducing human elements into mice, i.e., humanizing them, is to provide a more secure basis for extrapolating results onto humans (Maugeri and Blasimme 2011; Parkkinen 2017; Piotrowska 2013). That said, the process of humanization – model creation – is nevertheless conceptually distinct from model extrapolation because while the former simply concerns the intention of creating a model suited for a particular research task (i.e., it pertains to representational accuracy), the latter introduces the next step, namely the evaluation of the extent to which the newly created model succeeds in achieving the given research goal (i.e., it pertains to predictive accuracy).

3.4. Friends and foes of the similarity account of scientific representation

The concept of similarity has been discussed widely in the philosophical literature, with some authors defending some version of the similarity account of scientific representation (Giere 1988; Glennan 2017; Godfrey-Smith 2006; Khosrowi 2020; Mäki 2005; Teller 2001; Weisberg 2013), whereas others elaborate the concept in a specific disciplinary context (e.g., Sterrett 2017 who dicusses the concept

of physically similar systems). Ultimately, the analysis presented in this chapter attempts to shed some light on both the general discussion on representation and a specific disciplinary context.

In one form or another, the idea that a resemblance or a similarity relation would ground representation stretches far back into the past. In the context of the philosophy of science, the brightest spotlight was shone upon the similarity account by Ronald Giere, whose seminal work (Giere 1988) has served as a reference point since its publication. In the succinct words of Godfrey-Smith (2006, p. 726), "models are used to represent the world, via resemblance relations between the model and real-world target systems." However, resemblance or similarity is said to always come with "at least an implicit specification of relevant *respects* and *degrees"* (Giere 1988, p. 81)**.**

Despite its general acceptance and intuitive appeal, many philosophers working on scientific representation have challenged the account.[66] Arguably, one of the strongest arguments against the similarity account – and the only one that I address herein – is that representation is conceptually distinct from the notion of accurate, successful, or otherwise faithful representation. Frigg and Nguyen have argued that "representation is a wider concept than accurate representation and that representation cannot be analyzed in terms of accurate representation" (Frigg and Nguyen 2017, p. 54); Suárez emphasizes that "the puzzles regarding the notion of representation are prior to and independent of the issue of accuracy" (Suárez 2010, p. 93); others voice much the same sentiment (see especially Contessa 2007, p. 55 and 62; Kennedy 2012, p. 326). Crucially, similarity is explicitly taken to be an evaluative criterion of accuracy by at least some, if not all, of these authors. Indeed, as Frigg and Nguyen ([2017], p. 62) claim, "rather than being the relation that grounds representation, similarity should be considered as setting a standard of accuracy."

A distinction between representation and accurate representation is also maintained by some of those who are sympathetic towards the similarity account of representation. For example, Mäki takes representation as possessing "two major aspects: the representative aspect and the resemblance aspect" (Mäki 2005, p. 304). However, he further claims that "whether something is a representative of what it represents, whether it is a model as representative, is often revealed by whether it gives rise to questions or issues of resemblance" (Mäki 2005, p. 305). Finally, he adds that "considerations of resemblance presuppose that a system is employed as a representative, but on the other hand those considerations may serve as a criterion that helps identify a system as having the status of a representative" (Mäki 2005, p. 305). Thus, resemblance or similarity cannot in any straightforward way be kept separate from the notion of representation. Instead, there appears to be a reciprocal relation between similarity considerations and the establishment of a representational relation, i.e., something standing for something else. Mäki's account seems to fit well with what is going on in the research that makes use of mouse models. For instance, commenting on the advantages of humanized mice, Morton and colleagues state that:

[66] Within philosophy there is a long tradition of treating resemblance/similarity accounts with a high level of suspicion (e.g., Goodman 1976). Many of the contemporary objections have roots in Goodman (1976). It has been argued that the concept of similarity exhibits logical properties different from those of the concept of representation: whereas similarity is reflexive, symmetric and transitive, representation is neither (Frigg 2006; Frigg and Nguyen 2017; Suárez 2003, 2004). Furthermore, similarity is neither necessary nor sufficient for representation (Frigg and Nguyen 2017; Suárez 2003; Toon 2012a). Knuuttila (2005, 2011) argued that the similarity account inadequately views representation as a two-place relation between a model and its target, leaving out the crucial role of the scientist who does the representing. As much as these objections deserve full scrutiny, addressing them is beyond the scope of this chapter. Others have attempted to provide some answers, such as (Callender and Cohen 2006; Chakravartty 2010; Khosrowi 2020; Poznic 2016; Weisberg 2013).

> "Humanized mice (HM) allow researchers to examine xenograft growth in the context of a human immune system and resultant tumor microenvironment, and recent studies have highlighted the *increased similarities* in attendant tumor structure, metastasis, and signaling to those features in cancer patients. This setting also facilitates the examination of investigational cancer therapies, including new immunotherapies" (Morton et al. 2016, p. 6153, italics added for emphasis).
>
> "Many model systems either cannot propagate the disease in question or provide a foreign milieu, *not representative* of the conditions in humans. To address these challenges, chimeric systems designed to incorporate *relevant* human genes or tissues into a disease model organism have been developed" (Morton et al. 2016, p. 6153, italics added for emphasis).
>
> "These 'humanized mice' aim at harboring an immune environment capable of more *accurately reflecting* that present in human diseases" (Morton et al. 2016, p. 6153, italics added for emphasis).

I concur with both Mäki and the critics: similarity is intertwined with representation, and it is a criterion of accuracy. The apparent discrepancy is dissolved as soon as one considers the different epistemic roles of similarity considerations in the above three research modes.

In model selection, similarity considerations greatly influence whether a mouse model is chosen, i.e., whether the model is used as a representation. It is not necessarily the case that the more similar the model is to its target, the more likely it will be used, since model selection is also determined by a host of other factors. However, should the researchers reach the conclusion that effectively no relevant similarities arise, the model will be abandoned and no longer used as a representation (unless, of course, one studies the model as a negative model). In model creation, similarity considerations concern the targeted changes to be introduced into the model for it to serve as a representation. Therefore, in both model selection and model creation, similarity considerations play a major role in establishing and maintaining a representational relation. Consequently, the strict dichotomy between the notions of representation and accurate representation does not hold. However, this argument holds only for as long as one adopts an intentional approach toward representation, that is, for something to count as a representation it must be used as such. Many authors writing on scientific representation think that representation must, at least partially, be viewed in terms of the intentions of the scientists who make use of models to represent their targets (Giere 2010; Knuuttila 2011; Suárez 2010; Vorms 2011). If scientists no longer use certain mouse cohorts, these cohorts no longer count as representations.[67]

In contrast, in model extrapolation the similarity considerations pertain to the justification of the extrapolative inferences. Similarity in this sense concerns the evaluation of the accuracy of the results

[67] A seemingly radical answer to the problem of what constitutes a (scientific) representation is given by Callender and Cohen (2006), who argue that establishing a representational relation comes down to the act of stipulation: anything can serve as a representation of anything, provided that one so stipulates. However, they are careful to note that while some representational vehicles will be useful, others will not, and they further claim that "the questions about the utility of these representational vehicles are questions about the pragmatics of things that are representational vehicles, not questions about their representational status per se" (Callender and Cohen 2006, p. 75). Although many aspects of their account have been extensively criticized (Boesch 2017; Frigg and Nguyen 2017, pp. 55–57, 2020, pp. 23–30; Gelfert 2016, pp. 30–33; Morrison 2015, pp. 125–129; Toon 2012b, pp. 252–253), the fact that stipulation may play some part in establishing a representational relation remains largely undisputed. The argument of this section is that such stipulation is influenced by similarity considerations in selecting and in creating mouse models.

obtained by studying a given mouse model extrapolated to humans, i.e., predictive accuracy. Thus, if we were to consider this specific sense of similarity in isolation from the other research modes, it would suggest that similarity was merely an evaluative account rather than an account of representation. It is, however, also worth noting that model extrapolation presupposes that a model has been used as a representation of its target, which brings us back to the consideration of model selection or model creation, both of which are influenced by similarity considerations.

3.5. Chapter summary

The repertoire of mouse models used in cancer research and cancer immunology is vast. Given that there are different kinds of mouse models developed by numerous and diverse techniques, it should come as no surprise that each model has its own set of advantages and disadvantages. This chapter has analyzed the role(s) of similarity considerations in different research modes. The selection of a mouse model is guided by the research question at hand, a host of pragmatic and other factors, and, importantly, by similarity considerations. In model extrapolation, similarity concerns the evaluation of a mouse model and thus the justification of extrapolative inferences. Finally, in model creation, similarities pertain to the intention to actively introduce changes into mouse cohorts so that relevant similarities arise.

Clarifying these research modes and the role(s) of similarity considerations in the specific disciplinary context of mouse models of cancer also helps to shed some light on the debate on similarity in scientific representation. In particular, I have argued that whereas in model extrapolation the role of similarity suggests that a conceptual distinction between representation and accurate representation can be maintained by construing the latter in terms of predictive accuracy, it holds for neither model selection nor model creation. This is because in the two latter research modes, similarity considerations play a key role in the process of establishing and maintaining a representational relation.

4. Modeling epidemics and policy decision making: Analyzing an agent-based model of the SARS-CoV-2 epidemic

Because this chapter is based on joint work with Mariusz Maziarz, I will change the narrative by using the pronoun 'we'.

4.1. Introduction

In the aftermath of the outbreak of the novel coronavirus, governments around the globe have introduced non-pharmaceutical public health interventions aimed at slowing down the spread of the resultant pandemic. These measures range from relatively mild requirements like wearing face masks, washing hands, or avoiding close contacts to school closures and imposed isolation that are likely to have a detrimental and unpredictable influence on social and economic life (Wilder-Smith and Freedman 2020). Despite their significant impact, the introduction of many of these measures was not supported with high-quality evidence. First, conducting RCT would not be feasible for both ethical and practical constraints. Second, significant differences between the coronaviruses that caused the SARS and MERS outbreaks and SARS-CoV-2 (such as the likely airborne transmission (Lewis 2020) and asymptomatic infectiousness of the latter (Bai et al. 2020; Li et al. 2020)) undermine extrapolation from the data gathered during these previous epidemics. Finally, the current pandemic has not lasted long enough to gather observational data in the amount and quality sufficient for the assessment of the efficacy of alternative public health interventions, since the first reports were published just weeks after the first measures were introduced (Pan et al. 2020).

One of the many ways to address the issue concerning the impracticality of conducting RCTs and observational studies in the context of an ongoing pandemic is through scientific modeling, in particular epidemiological modeling. Here, we focus on the so-called agent-based modeling (ABM) approach, which differs from more traditional epidemiological modeling in several ways.

ABMs are a form of computational modeling strategy where agents are treated as entities interacting with each other and their environment in a locally-defined fashion described by a set of rules. The overall dynamics of the system are then computed, allowing for the simulation of complex patterns and an understanding of how these patterns arise (Railsback and Grimm 2012; Wilensky and Rand 2015). ABMs are used in many scientific contexts, including modeling the spread of infectious diseases, and have proven successful in informing policy decisions before. For instance, Eisinger and Thulke (2008) modified and then applied a previously-developed ABM of the spread of rabies, generating a rule-based model that represented specific spatial and behavioral characteristics of the fox population: e.g., fox families represented as moving within home ranges and young foxes engaging in long-distance migratory behavior (Railsback and Grimm 2012). Whereas the classical differential equation models predicted that vaccinating at least 70% of the fox population would eliminate rabies, the ABM indicated that a successful vaccination strategy could do with much less than 70% of the population being immunized once the spatial arrangements of fox hosts were explicitly considered, saving millions of Euros as a result. Moreover, the ABM also suggested that the classical strategy would fail more often than not, and it was successfully applied to deal with the rabies problem. However, despite the promising record of using ABMs in effective epidemiological interventions, its use in informing proposed measures against the novel coronavirus epidemic has raised criticism (Ferguson et al. 2020; Squazzoni et al. 2020; Sridhar and Majumder 2020).

Unfortunately, for the assessment of healthcare interventions based on this type of epidemiological models, standard evidence hierarchies exclude agent-based models altogether and include theoretical or mechanistic inferences at the lowest level of the hierarchy. For example, the Oxford Centre for Evidence-Based Medicine (OCEBM Levels of Evidence 2009) and the National Institute for Health and Care Excellence (NICE guidelines) (National Institute for Health and Care Excellence 2014) include theoretical and mechanistic reasoning but agent-based models fall beyond their scope. This can be explained by the novelty of agent-based modeling and the limited trust of the evidence-based medicine (EBM) movement in theoretical and, to some extent, also mechanistic reasoning, which, despite being used implicitly to assess the possibility of confounding and the quality of results (Rocca 2018), is downgraded or rejected as either subjective or fallacious (Worrall 2010). However, such a view has been challenged by a group of philosophers advocating for improving the practices of evidence assessment in medicine by putting more weight on mechanistic reasoning in causal inference (Clarke et al. 2013; Parkkinen et al. 2018; Williamson 2019). The position of the EBM+ program (Clarke et al. 2013; Parkkinen et al. 2018; Williamson 2019) is encapsulated by the normative reading of the Russo-Williamson Thesis (Russo and Williamson 2007) which states that causal claims should be based on both difference-making and mechanistic evidence.

The causal claims supported by agent-based models have been interpreted inconsonantly: either as being in line with the potential outcome approach (POA) (Marshall and Galea 2015) as delivering theory-driven understanding (Hernan 2015), or as providing mechanistic evidence (Clarke et al. 2014). Below, we show that all of these apparently inconsistent interpretations are correct, because the best contemporary ABMs bear a resemblance to the actual mechanisms and therefore allow for the counterfactual assessment of intervention efficacy in the target while also delivering an understanding of the phenomenon of interest. Our argument proceeds by discussing as a case study an ABM of SARS-CoV-2 epidemic in Australia (Section 4.2.). We argue that the best ABMs represent actual mechanisms despite the presence of various simplifications (Section 4.3.). Finally, we consider the limitations of using ABMs as evidence for policy decisions (Section 4.4.).

4.2. Modeling the SARS-CoV-2 epidemic

Apart from the compartmental SIR (Susceptible, Infectious, Recovered) framework and its derivatives (Acuna-Zegarra et al. 2020; Giordano et al. 2020; Kissler et al. 2020; Neher et al. 2020; Peng et al. 2020; Yang et al. 2020) and regression analysis (Fu et al. 2020; Tobías 2020), most advanced models of the spread of the novel coronavirus are transformed versions of agent-based influenza pandemic models (Chang et al. 2020; Ferguson et al. 2020). Such models have been used as evidence for introducing (sometimes severe) public health measures (Adam 2020), with the recent change in British policy being the prime example. In this section, we illustrate this approach to modeling the SARS-CoV-2 pandemic with an agent-based model of the epidemic in Australia (Chang et al. 2020) based on ACEMod (Australian Census-based Epidemic Model). Developed as a "framework for studying influenza pandemics in Australia" (Cliff et al. 2018, p. 412), ACEMod is an influenza spread model that addresses the need for simulating interventions responding to the outbreaks of future respiratory diseases. While the 2009 swine flu pandemic was the motivation for constructing ACEMod, the model was not intended to accurately represent the outbreak of the H1N1 strain, but rather as a generalized framework for studying how an infectious disease spreads through the social interactions of Australians. ACEMod utilizes census data to ascribe realistic spatial and social characteristics to almost 20 million agents inhabiting the model world. These agents are divided into different social groups of varying characteristics, with households differentiated proportionally according to statistical data on the prevalence of different types of families (singles, single parents, and couples with or without

children). These features are ascribed to agents stochastically in a way that replicates the aggregate structure of statistical data. During the daytime, children and students meet in classrooms and at schools, adults go to work, and pensioners stay at home. During the nighttime, the agents encounter contacts at households and in their neighborhoods (e.g., at supermarkets, theaters).

The disease can be contracted by an agent in the event of meeting an infected individual in one of these settings. The probability that an agent *i* contracts the disease in a given step *t* depends on the number of sick individuals met in that step and the contagiousness of the disease, scaled by *K.* The modelers assume that the infectivity of the disease decreases linearly over time. Asymptomatic cases are assumed to be 50% less infectious than symptomatic ones, and the flu lasts 5 days within the model. After this period, recovered agents cannot infect others. Additionally, those who experience symptoms do so after an incubation period lasting approximately three days. The influenza epidemic is started by agents coming to Australia via international airports and seeded into communities living near the airports at random.

In order to represent an epidemic of a particular strain of influenza using ACEMod, the model requires calibration. Modelers can proceed with this step in two ways, depending on the accessibility of data. In the case of well-studied influenza strains, their infectivity and the ratios of transmission in different contexts are well-recognized, and parameter values can be chosen on the basis of empirical studies. However, if these data are missing, then parameter values have to be calibrated using statistical procedures such as simplex or genetic algorithms to maximize the fit of the model to a benchmark. After constructing and calibrating ACEMod, modelers run simulations to obtain the estimates of prevalence, incidence, and attack rates, and choose the most common outcome (due to stochasticity, different runs of the model may lead to obtaining slightly different results).

Chang et al. (2020) have used a significantly amended version of ACEMod to address the question of the effectiveness of non-pharmaceutical interventions aimed at suppressing the SARS-CoV-2 epidemic in Australia. The selection of models constructed to control a novel and possibly deadly strain of the seasonal flu in this case is primarily the result of the rapid demand for evidence informing decisions regarding public health measures, which may raise doubts about the justification and soundness of their conclusions. For example, one can ask whether the efficacy claims assess healthcare interventions against the novel coronavirus epidemic or an artificial pathogen existing only within the model world that shares some features of influenza and others of SARS-CoV-2. To address this criticism (considered in depth below), we discuss the changes introduced to the model and argue that the process of model calibration and validation suggests that the model represents the actual mechanism of the SARS-CoV-2 epidemic.

ABMs such as ACEMod can be seen as consisting of two parts: the rules specifying the behavior of agents and the creation of the model society, as well as the assumptions characterizing the infectivity of the pathogen causing the epidemic. Given that ACEMod is based on 2016 census data and a major change in social behaviors is unlikely to have occurred since then, the model accurately represents the social interactions of present-day Australians. Hence, the former part of the model has been left mostly unchanged, beyond increasing the number of agents to over 24 million to adjust for the growing population. In addition to introducing a social structure sufficiently resembling the contact network of the present population, obtaining accurate predictions of epidemic development and policy assessment requires inputting data on transmission likelihoods that are true for the pathogen causing the modeled epidemic (Cauchemez et al. 2011). Most changes in the model are concerned with the assumptions specifying the infectivity of the disease. Even though several features of influenza epidemics are similar to the epidemic caused by the novel coronavirus, they differ with respect to infectivity and attack rates, mortality rates, the average duration of disease, the reproductive number

R_0, and the distribution of asymptomatic cases. Therefore, these parameters in the model required recalibration.

The transmission probabilities remained mainly as specified in the influenza model. In order to account for the differences in the incubation period and disease length, Chang et al. set the time from contraction to the appearance of symptoms to 5 days on average and the duration of the disease to 12 days. Infectivity increases exponentially the day after an agent gets infected and then decreases linearly until the end of infection, so cases are most infectious at the start of symptoms. The length of the generation period was calibrated to 6.4 days in order to reflect this difference in the model. Additionally, the likelihood of contracting SARS-CoV-2 but staying asymptomatic was set to be age-dependent, and equaled 1/3 for adults while minors were set to be five times less likely to suffer from symptoms than adults. While this assumption is in agreement with the empirical findings that children represent a minor fraction of symptomatic cases, the calibration aimed at reproducing aggregate epidemic curves and may diverge from the actual chances of developing symptoms.

Within the ACEMod framework, the reproductive number R_0 is not one of the assumptions inputted into the model. Rather, its estimate results from a simulation of the scenario described by the rules and assumptions, some of which are stochastic. The assumptions considered and, particularly, the parameter denoting contagiousness of the disease (K) have been calibrated such that R_0 stays within the limit of (2.0-2.5), i.e., in agreement with empirical estimates of the reproductive number at the beginning of the SARS-CoV-2 outbreak (Lai et al. 2020; Liu et al. 2020). The set of parameter values that result in the estimate of R_0=2.27 create the epidemic dynamics reproducing the beginnings of the outbreak in a few countries experiencing the disease prior to Australia (China, Italy, Spain), where the growth rate of cumulative incidence equaled roughly 0.2. In addition to reproducing the empirical data for the beginning of the epidemic, the recalibrated ACEMod allows for simulating what the future of the epidemic in Australia may look like. As the modelers admit, the *Baseline* scenario, which is based on the assumption that agents do not change their behavior in response to the epidemic, is unlikely given the widespread self-imposed isolation in other countries. However, it allows for counterfactual comparisons of the different possible (sets of) interventions relative to the Baseline scenario. In order to assess the efficacy of particular healthcare policies, Chang et al. modify relevant rules and assumptions to describe the spread of SARS-CoV-2 under case isolation, school closure, three levels of compliance with social distancing, and with a few combinations of the three policies. For instance, in order to assess the effect of school closure (including primary and secondary schools, colleges, and universities), the parameter denoting the chance of meeting an infected agent in schools is set to zero, which describes the situation when both students and teachers stay at home (and hence cannot contract the virus). These counterfactual scenarios represent the effects of interventions on the model world. All interventions are modeled as taking place after the number of cases exceeds 1000. The comparison of most common outcomes (given the stochasticity of the assumptions and rules, they are indeterministic) including interventions with the baseline scenario allows for putting forward counterfactual causal claims that describe the effects of interventions on peak incidence and prevalence and the development of the epidemic in time. The conclusions accurately describe the effects of interventions within the model as long as no coding error occurs. However, the reliance of the model on simplifications generates a question as to whether the assessment of intervention efficacy holds for the novel coronavirus epidemic in Australia.

4.3. ABMs as models of actual mechanisms

Before proceeding to our argument, let us first make several general remarks about modeling. These remarks should prove essential in clarifying the main issues that are often raised with regard to using simplified models, particularly in the context of policy decision-making. First of all, ABMs are instances of mechanistic models, for they clearly fit the general, also called the minimal, characterization of what a mechanism is: a set of entities whose activities and interactions are organized such that they are responsible for the phenomenon (Glennan 2017; Glennan and Illari 2018; Illari and Williamson 2012). This definition is broad enough to conceptually unify the debates on biological and social mechanisms under a single notion of a mechanism. Furthermore, such definition leaves open the possibility of integrating biological and social aspects into a mixed-mechanism model (Kelly et al. 2014).

It should also be noted that much like any other kind of model, ABMs serve as simplified representations of their target phenomena. As the ACEMod case clearly shows, modelers introduce various simplifications by which they purport to adequately capture the core dynamics of the modeled phenomenon. In this process, they first abstract away from the complexities of the real system by 'extracting' certain features that they believe to be of crucial importance and that will then be the focus of modeling, whereas other features that may or may not have a causal influence are disregarded in these early stages. Modeling is an iterative process during which the merits of the model's assumptions are continuously being evaluated, and if required, the assumptions are refined and additional assumptions added. More importantly, some of those extracted features are distorted to the extent that, if taken literally, they would misrepresent the actual state of things. However, introducing such distortions is often made in full awareness, with the ultimate goal of finding out whether the consequences they have for the behavior of the system make a difference, and to what degree. Philosophers often refer to the former (i.e., the set of properties retained in a model) as an abstraction, while the latter (i.e., the distortions of the system's features) is called an idealization (see Chapter 2).

However, abstractions and idealizations do not exhaust the conceptual toolbox available to modelers. A popular way to attempt to model a given system realistically is to introduce various approximations. Although there are noteworthy differences between approximations and idealizations, we cannot afford to go into any detail here. In summary, models often effectively disregard, distort and otherwise simplify possibly important details. In light of this, many wonder whether we can gain insight into the modeled phenomenon at all, and if so then how.

Although the SARS-CoV-2 ABM is fairly detailed and precise, it cannot do without some of the simplifications discussed above. Consider some of the following assumptions introduced in the model. On the one hand, the basic features of the social life of the majority of the population are extracted and considered in the model: e.g., the inclusion of day and night regimes with their respective differences in social behavior allows for modeling a more realistic scenario than in simpler models. On the other hand, the infectivity of symptomatic and asymptomatic cases is considered to be constant for all members of the two groups of agents, albeit it differs between the groups. In reality, we expect that infectivity varies, which is further supported by extreme cases of super-spreaders who infect a large number of people and thus may seed new local outbreaks, which could arguably impact the predictions (Frieden and Lee 2020; Lloyd-Smith et al. 2005; Wong et al. 2015). Other parameter values also have a wide distribution but are treated as constant, often by calculating the mean value. The ABM also does not consider the potential impact of ethnic differences (Delgado et al. 2002; Everhart et al. 2000; Lazarus et al. 2002; Redelings et al. 2007; Smith and Clatworthy 2010) in the population with respect to different lifestyles, socioeconomical status and immune host responses, all of which could affect the dynamics of the spread.

Furthermore, some other assumptions exceed our current understanding of the epidemic and SARS-CoV-2's transmission mechanism. For example, one of the assumptions of the ACEMod model is the linear reduction of infectivity over time. Unfortunately, empirical results (Zou et al. 2020) suggest only that infectivity reduces over time, but do not indicate the linearity of this process. Additionally, ACEMod and its SARS-CoV-2 version put agents into working groups of 20 agents, despite the heterogeneity of their working conditions. Considering the differentiation of work duties (from healthcare workers and shop assistants to writers with virtually no social interaction), the chance of meeting an infected person at work is actually job-specific and therefore the model simplifies the reality.

Consequently, we concur with Andersen's claim that "no mechanism model can include all the actual, much less the potential, causal relationships in which such a mechanism may engage in a system" (Andersen 2012, p. 995). This pessimistic view on simplified models has inspired the method known as exploratory modeling (Bankes 1993). In cases when the values of parameters and assumptions inputted into the model cannot be established with certainty, researchers can simulate multiple possible worlds to discover the dependencies that are stable across the set of different models. In cases when only a fraction of assumptions are uncertain, researchers conduct sensitivity analyses to check if changes in the values of the parameters lead to changes in their conclusions (Wu et al. 2013). The results that remain unchanged despite minor adjustments to assumptions are considered to be robust (Weisberg 2006). This, in turn, leads to choosing those interventions that are most effective across different sets of parameter values, known as robust decision-making (Bankes 1993).

Others prefer to think in terms of the distinction between how-actually and how-possibly modeling, referring to models that describe an actual mechanism or a possible mechanism, respectively (Machamer et al. 2000). There are two general ways to unpack the concept of a how-possibly model. First, we may want to say that a model serves as a hypothesis to be confirmed or disconfirmed as new evidence emerges. In this sense, a how-possibly model will eventually either turn into a how-actually model, should the evidence confirm it, or be discarded if the evidence is contrary to the model's conclusions. The other general notion of a how-possibly model invites a different attitude. Rather than being in the position of having little data to establish whether or not the model does, in fact, represent the actual mechanism, we may interpret the model as representing something other than the potentially actual mechanism. On this view, claims about possible mechanisms do not attempt to pick out actual states, nor do they attempt to explain how a phenomenon actually occurs. Instead, they refer to conceivable states, and ask whether the hypothesized mechanism could, in principle, produce the phenomenon in question if certain assumptions are satisfied.

Here we argue that, notwithstanding the simplifications introduced in the discussed influenza and SARS-CoV-2 ABMs, the epidemiologists are, in fact, providing representations of actual mechanisms of the spread of the viruses. This can be supported by exploiting the relevant similarities (Giere 2004, 2010) between the SARS-CoV-2 ABM and the actual outbreak. The respects in which an ABM can be judged similar to its target concern the features retained in that model, while the degree(s) of similarity concern the extent to which the model's features match those of the phenomenon. A good example is calibrating the incubation period to 5 days, based on existing studies according to which "the mean incubation period was reported as 5.2 days, 95% CI [4.1, 7.0], while being distributed around a mean of ~5 days within the range of 2–14 days with 95% CI" (Chang et al. 2020, p. 8).

To elaborate this further, we may draw on Glennan (2005) who introduced a useful conceptual distinction between what he calls behavioral adequacy and mechanical adequacy. According to Glennan, a model represents an actual mechanism if it reproduces the aggregate behavior of the phenomenon, and truthfully describes its parts and interactions. Concerning the behavioral adequacy, one should be asking if "the model predict[s] (quantitatively or qualitatively) the overall behavior of

the mechanism" (Glennan 2005, p. 457). By calibrating the model to data from the beginning of the epidemic, Chang et al. (2020) showed that it reproduces the benchmark variables (R_0 and attack rate).

Two remarks are in order here. First, one may oppose the claim that what is being represented is the actual mechanism by arguing that the mechanism underlying the beginning of the outbreak and the fully-fledged epidemic are distinct. Changes in social behavior or genetic mutations could undermine the behavioral adequacy of the model. Second, it is possible (at least in principle) that the model represents a false mechanism, but is calibrated to the relevant benchmark such that it reproduces it. For example, there is no data confirming (or disproving) the assumption that children are asymptomatic five times more often than adults. As the modelers admit, this assumption was made not only to account for the lower attack rate among minors, but also to make the model adequate to aggregate-level data. This approach to calibration resembles the estimation of statistical parameters (a.k.a. curve fitting) and is considered dubious. The main line of criticism highlights that it is in principle possible to construct a model that represents a possible mechanism and, using calibration, adjust parameter values so that it reproduces the represented phenomenon, i.e., obtains behavioral adequacy despite being false. However, while this criticism is indeed justified regarding models of mechanisms that are epistemically inaccessible in other ways (such as mechanisms in the social sciences (Maziarz 2020)), it is not so in the case of epidemiological mechanisms whose transmission mechanism can be studied empirically and compared to the mechanism represented by the model.

This can establish that the mechanism represented by the model is similar (in relevant aspects and to relevant degrees) to the mechanism that generates the outbreak, i.e., achieves mechanical adequacy in Glennan's terminology. Applying the list of Glennan's (2005, p. 457) criteria for mechanical adequacy justifies the claim that the mechanism represented by Chang et al. (2020) resembles the actual mechanism. First, according to our best contemporary understanding of the spread of the novel coronavirus, the model identifies all of the components of the mechanism. This would change if further studies identified other significant transmission routes, e.g., the fecal-oral route. Second, the model represents the entities of the mechanism in a localized way, given that it retains the spatial distribution of inhabitation in Australia. Additionally, the model simulates the development of an epidemic in time. This asserts that the "spatial and temporal organization of the mechanism" is accurately represented. Third, given that the number and place of social interactions are crucial for modeling the spread of contagious diseases, the model accurately captures relevant properties of the agents inhabiting the model world. Fourth, the calibration to census data asserts that the model provides "quantitatively accurate descriptions of the interactions and activities of each component," at least on average for groups of agents. Finally, our background knowledge suggests that there is no other mechanism (different from the spread of the pathogen through human interactions) that could be responsible for the epidemic of SARS-CoV-2.

Given that ACEMod fulfills Glennan's criteria for behavioral and mechanical adequacy, considering our current understanding of the novel coronavirus, we can conclude that Chang et al.'s (2020) model represents the actual mechanism of the spread of the disease in Australia. Given this, the claims assessing the efficacy of the mitigation measures under consideration are likely to be accurate not only within the model but also about its target. We claim this with several caveats in mind to be discussed in the next section.

It is also important to note that the ABM integrates the biological aspects, expressed by the parameter of infectivity, and the social aspects such as daily interaction regimes. As a result, the ABM should be construed as an instance of a model of a mixed mechanism, a concept elaborated by Kelly et al. (2014). Due to exposure patterns, population-level phenomena such as infectious disease epidemics are crucially dependent on human behavior and social practices. In cases like the current pandemic,

effective interventions may best be aimed at the societal level and therefore mechanistic models that integrate social factors, human behavior and biological aspects (something that the ABM discussed here attempts to do) are arguably best suited for providing understanding and suggesting policy decisions.

4.4. Discussion and recommendations

Our study defends using ABMs for informing decisions regarding mitigation and suppression measures by arguing that its best epidemiological models represent actual mechanisms. Provided that the model's assumptions are calibrated and checked against the background empirical data - that is, the components, their activities, and spatiotemporal organization resemble (in relevant aspects and to a certain degree) the actual state of things - iterative runs of the simulations can indeed provide understanding and inform policy decisions. This is because the model delivers both difference-making and mechanistic evidence by satisfying the criteria of behavioral and mechanical adequacy, respectively.

In contrast to our claim, epidemiological SIR models and ABMs have been criticized for over-simplifying target phenomena and hence lacking relevance for policy decisions. For instance, Eubank et al. criticized the Imperial College London model (Ferguson et al. 2020) for its "reliance on a simplified picture of social interactions [that] limits its extensibility to counterfactuals. The general nature of conclusions based on such model can be expected to be similar to those of a simple compartmental model" (Eubank et al. 2020, pp. 5–6). Similarly, Squazzoni et al. suggested that even though ACEMod is better calibrated than other epidemiological ABMs, "these [models] do not capture network effects nor people's reactive responses as the population states simply change via stochastic (randomized) processes determined by parameters (although the parameters derive from data)" (Squazzoni et al. 2020, p. 2.6). In our view, these highly-advanced epidemiological models, while being simplified representations of reality, account for relevant aspects of social interactions and crucial aspects of the novel coronavirus epidemic (e.g., contagiousness), therefore allowing them to be put forward as evidence for policy-relevant claims.

We claim this despite that a straightforward comparison of model predictions to the actual epidemic curve (e.g., the number of total cases) in Australia shows the two to be mismatched. The number of COVID-19 cases is smaller than predicted by an order of magnitude. However, such a direct comparison is not warranted because the countermeasures implemented by the National Cabinet and the state governments differ from the mitigation and suppression interventions considered by Chang et al (2020). That is, the *a posteriori* behavioral adequacy of the model cannot be directly assessed based on the predictions because the scenarios implemented into the model differ from the actual course of events. In particular, first restrictions on international travel were imposed on March 1st, when just 29 COVID-19 cases were observed ("Worldometer" 2020), followed by the 14-day quarantine for incomers (Pannett 2020) on March 15th (300 cases) ("Worldometer" 2020) that virtually stopped the import of new cases to Australia, the closure of borders for nonresidents (Worthington and Snape 2020), and a social distancing rule requiring 4 sq. meters for each person in enclosed space on March 20th (928 cases) ("Worldometer" 2020). Two days later (1 609 cases) ("Worldometer" 2020) some states closed non-essential businesses (Knaus et al. 2020), and on March 30th (4 460 cases) ("Worldometer" 2020) they forbade gatherings of more than two people and advised people to stay at home with some exceptions ("National Cabinet Statement" 2020). The last two interventions are more severe than the measures considered by the modelers and are a plausible explanation of the overestimation of the number of cases. Given this, we can claim that the model had been behaviorally

adequate to the mechanism governing the beginning of the epidemic in Australia and it would have produced accurate predictions had the interventions been introduced in line with the measures simulated by Chang et al. (2020). However, inaccurate predictions are what should be expected in the case of the so-called fat-tail processes, where outcomes strongly depend on the initial conditions. One should expect that, over time, the assumptions and calibrated parameters will be more accurate and ABMs will produce predictions not only qualitatively but also quantitatively accurate. The usefulness of epidemiological ABMs for decision-makers results from delivering an understanding of the spread of the virus and allowing for comparisons among alternative mitigation measures. For instance, one of the qualitative predictions of the model is the limited efficacy of school closures, which remained open in Australia (Karp and Davey 2020) and which had limited influence on the severity of the epidemic, considering that just one cluster was located at a school ("Coronavirus update for Victoria" 2020).

We believe that, considering the diversity in the number and patterns of social interactions across countries, the quality of evidence from ABMs should be assessed on the case-by-case basis. To do so, one can employ the approach of Parkinnen et al. (2018, p. 79) developed initially to evaluate the quality of evidence for biological mechanisms. In that case, one should consider (1) the quality of the method (i.e., consider the empirical adequacy of the assumptions in light of contemporary empirical results), (2) the implementation of the method (i.e., assess how the epidemiological ABM is programmed, calibrated, and simulated), and (3) the stability of the results (i.e., how sensitive the results are to changes in the assumptions). ACEMod (Chang et al. 2020) fulfills these criteria (provisionally accepting the existing empirical results but keeping in mind that they may change as the pandemic develops in time and new results become published).

Epidemiological models usually do not account for the harms of non-pharmaceutical interventions. Severe mitigation measures such as imposed social distancing and business closures are likely to hamper economic and social life. All models are partial representations of reality and, given that the primary purpose of an epidemiological model is to address the efficacy of health care interventions, they isolate away certain factors and effects of interventions (economic and social) and are more accurate in predicting the spread of the disease under alternative conditions. Other models (Bodenstein et al. 2020; Dignum et al. 2020) trade off epidemiological accuracy for accounting for social and economic effects, and may be more relevant for assessing the harms of mitigation measures.

Additionally, ABMs, much like the compartmental models, are dependent on the assumptions of the modelers (Sridhar and Majumder 2020). Our claim that ACEMod calibrated for SARS-CoV-2 bears similarity to the actual mechanism of the epidemic, depends on the accuracy of the empirical results used as an input for this model. We need to repeatedly acknowledge the provisional nature of these empirical results, given the novelty of the pathogen. If the parameter values in ACEMod were miscalibrated, then the assessments of intervention efficacy could be wrong. This implies that problems arise when the features of the virus change due to mutations and when people change their behavior in a significant and unpredictable way since "the efficacy of implementation depends on people's reactions, [the stability of] pre-existing social norms, and structural societal constraints" (Squazzoni et al. 2020). Furthermore, the effects of epidemiological agent-based modeling are highly dependent on social structure and carefully calibrated to social and economic characteristics. Therefore, the epidemiological ABMs are geographically-localized and their conclusions should not be extrapolated beyond their target systems (Broadbent and Smart 2020), unless the models and their predictions are calibrated to particular settings. Finally, while ACEMod is well-documented in the two publications discussed throughout our paper, neither its code nor detailed documentation regarding its use is published (this unfortunately also applies to some other ABMs of the SARS-CoV-2 epidemic).

Given these limitations, the models should be carefully checked for coding errors and other possible flaws before applying their implications in the policy context.

Finally, an additional concern which builds on much of the previous discussion must be addressed. The EBM movement assesses the quality of evidence on the basis of considering the risk of bias or confounding, i.e., the situation where a variable left out from the model accounts for estimated correlations and is the actual causal factor (La Caze 2009). The methods that are less likely to lead to spurious correlations are considered more trustworthy. This approach prioritizes randomized controlled trials over observational studies and cohort over case-control design. Considering the crucial role of epidemiological ABMs in planning mitigation and suppression measures during the SARS-CoV-2 pandemic, the question of the quality of evidence delivered by such models needs to be assessed with regard to the problem of confounding.

As noted above, epidemiological ABMs have been criticized for simplifying social interactions and not accounting for changes in behavior in response to the pandemic and measures themselves (Squazzoni et al. 2020). Much like other scientific practices, theoretical modeling consists of a careful selection of factors that are considered relevant for the task at hand. Thus, by extracting certain features of the studied phenomenon, which, in many cases, are then further modified in various ways (Portides 2018), modelers can indirectly investigate the behavior of a system by first investigating their model (see also Chapter 1 and Chapter 2). In the process, however, it might be the case that some difference-making factors have been overlooked or otherwise misrepresented in a way that puts the utility of the model in question. Introducing such simplifications can result in findings that may be affected by a confounding factor that is not accounted for in the model (Strevens 2008, p. 288).

Consider, for instance, the attempt to assess the relative accuracy of assumptions in the SARS-CoV-2 version of the ACEMod model. The question remains whether some of the extracted factors, presumed to be key difference-makers, do, in fact, make a difference and whether they do so in accord with the assumptions. Chang et al. (2020) accounted for, among other things, age-dependent attack rates, a range of reproductive numbers, age-stratified and social context-dependent transmission rates, household clusters and other social mixing contexts, symptomatic-asymptomatic distinction, and long and varying incubation periods. However, the question whether all the key difference-making factors have been included in the model remains to be addressed.

For example, the modelers assumed temporal homogeneity of interactions among agents populating ACEMod. ACEMod predicts that SARS-CoV-2 spreads more quickly in urban areas because social contacts are more frequent in such settings. However, as shown elsewhere (Nation et al. 2010) these social interactions are less intensive in terms of duration than social interactions in rural areas, which suggests that the spread may be equally rapid or even faster in rural areas, depending on the strength of the confounder's influence. This is because the duration of exposure is a plausible moderator for the risk of contracting SARS-CoV-2, as stated by, e.g., ECDC (Adlhoch et al. 2020). The failure to properly account for the duration of exposure can have detrimental effects on the accuracy of predictions.

There are many potential confounding factors like the one just discussed that are not included in the model. Given the utmost importance of providing reliable predictions and accurate assessment of public health (non-pharmaceutical) interventions, further research is needed to assess the risk of confounding and quality of evidence delivered by epidemiological ABMs. This task can be divided into two intertwined questions. First, it needs to be established that all the key difference-making factors are accurately included in the model (i.e., no confounding factor is left out from the model). Second, all factors included in the model should be shown to make a difference (approximately) in accord with the rules and assumptions of the model so that the ABM does not include spurious determinants and

does not misrepresent the actual factors in a way that would lead to false predictions. Such analysis allows for addressing the question if the causal structure represented by an ABM resembles the actual causal structure and, in effect, assessing the quality and reliability of evidence. However, considering that epidemiological ABMs account for not only biological determinants such as infectivity but also social interactions that differ across the globe, the quality of evidence from ABMs must be assessed on a case-by-case basis. In reaching policy decisions, ABMs should be understood as merely one piece of the puzzle subject to further re-evaluation with respect to value judgments. This is because alternative mitigation measures may disproportionately affect certain social groups. Therefore, the quality assessment aimed at identifying possible confounders that have been left out of a particular ABM should delineate the conflict of interest and vested values related to the ABM and the mitigation measures that it supports.

4.5. Chapter summary

In summary, we have argued that, despite the criticism raised against models being the appropriate vehicle for informing policies, the SARS-CoV-2 ABM is suitable for this purpose because the mechanism described by the model sufficiently resembles the mechanism at work in the real world. Thus, our best contemporary epidemiological ABMs represent the actual mechanism of the spread of the virus. Unfortunately, such models have often been left out from methodological discussions and are not explicitly listed by evidence hierarchies. While the need for appraising mechanistic reasoning in medicine is also voiced by the EBM movement (Anjum et al. 2020), there is no broadly-accepted view on how to amalgamate evidence of different types. Further research is needed to assess the risk of bias and confounding in the epidemiological models that deliver both difference-making and mechanistic evidence. However, considering the current situation and pressing need for rapid and accurate decisions regarding mitigation measures, policymakers should take to heart the advice that "if no randomized trial has been carried out […], we must follow the trail to the next best external evidence and work from there" (Sackett et al. 2000, p. 74). In the current situation, accurately calibrated epidemiological ABMs are the best existing evidence.

5. A conceptual model of the immune system: Understanding immunity in times of COVID-19 and beyond

Because this chapter is based on joint work with Gregor Greslehner, I will continue using the pronoun 'we' as in the previous chapter.

5.1. Introduction

In their reflections on the COVID-19 pandemic, philosophers have shed light on a number of different aspects. We shall focus on a central aspect that has not yet received much attention: immunology. Philosophy of immunology has only recently started to grow as a small field within philosophy of science (see Pradeu 2019; Swiatczak and Tauber 2020). Immunology can be an overwhelmingly complicated science, even for experts who have worked in the field for decades. However, the basic principles and underlying theoretical concepts are also a domain for philosophical reflections that benefit immunology, philosophy, and the consideration of how a wider audience of non-immunologists think about immunology.

The COVID-19 pandemic caused by the spread of SARS-CoV-2 naturally invites talk of a host defense against a foreign invader, a pathogen, giving rise to the idea that the stronger the defense against the pathogen (the foreign 'non-self'), the better for the host (the 'self'). This idea is further illustrated by the benefits of boosting one's immunity by vaccination, or the communication coming from some health agencies such as the CDC (2021) stating that immunocompromised individuals possess weakened immunity, which is a risk factor. However, we shall argue that such a construal of immunity contributes to the distortion of the overall picture of *what the immune system is, what it does, and why it sometimes fails*.

Many features of COVID-19 painfully remind us of several issues concerning the immune system and raise important questions, including some which extend beyond COVID-19. These issues and questions include, but are not limited to: the contextuality of the immune response; the trade-off between fighting off an infection (immunity) and in so doing causing collateral damage (immunopathology); the two defense strategies, i.e., clearing the pathogen (resistance) and decreasing the susceptibility of the host to tissue damage (disease tolerance); the importance of immune regulation; and questions going well beyond the narrow conception of immunity as a defense system.

In this chapter, we propose a conceptual model of the immune system consisting of three features: contextuality, regulation, and trade-offs. This tripartite view allows us to take a broader perspective on many of aspects of COVID-19 and achieve a better understanding of the immune system in general. Using this model, we also want to draw attention to misleading metaphors originating from the idea that the immune system is primarily a defense system to fight pathogens.

War-like metaphors, such as defending the 'self' against pathogenic invaders, continue to shape how many scientists and physicians think about the immune system and how immunology is being communicated to the wider public.[68] Since metaphors have their uses and abuses, it is important to

[68] Even the best textbooks (e.g., Murphy and Weaver 2017) focus narrowly on defense at the expense of other immune functions, thus introducing a biased mindset in future generations of biomedical researchers and physicians.

see how they guide one's intuitions and how we think about the immune system. Using our model framework, we provide a non-exhaustive categorization of what the otherwise ill-defined notions of 'strong' or 'weak' immunity might mean, and we argue that some of the (outdated or questionable) distinctions and metaphors of self, danger, defense, and strength of the immune system or response, have led to misconceptions, limiting our understanding of the immune system.

As a result, we suggest that we need to move away from viewing the immune system narrowly as a defense system, and to drop related notions that prevent us from achieving adequate understanding. The conceptual clarification of these matters showcases the use of philosophy of science in the quest for a better understanding of the immune system.

Before moving on, it is worth noting that the use of the term 'model' does not correspond to either of the modeling practices discussed in Chapter 1. However, this does not mean that such use results in a blatant confusion of terms, for the topic of Chapter 1 was focused on the *modeling process*, rather than the *product*. Moreover, the product of the modeling process in Chapter 1 concerned mechanistic models while in this instance it pertains to a conceptual account of a whole system, rather than a particular mechanism underlying a biological phenomenon. Still, the use of 'model' herein is quite loose and akin to, for example, the equilibrium model of immunity proposed by Gérard Eberl (2016).

5.2. Rethinking immunity: what does the immune system recognize and respond to and what does it do?

The guiding principle that the immune system distinguishes between 'self' and 'non-self' has been dominant since Sir Frank Macfarlane Burnet (Burnet 1969; Tauber and Podolsky 1994). According to this view, anything belonging to an organism's self would be tolerated, whereas anything foreign would be recognized as non-self and removed or attacked. This basic idea has helped to explain various immunological phenomena, while many other problems and limitations have grown ever more apparent (Pradeu 2012; Tauber 2017). For example, it remained unclear how the immune system would be able to make this distinction. A major breakthrough came with Charles Janeway's and Ruslan Medzhitov's insight regarding the interplay of adaptive and innate immunity and the discovery of pattern-recognition receptors (Medzhitov and Janeway 2002). Suddenly it seemed as if everything that was foreign or pathogenic and needed to be attacked was wearing a molecular pattern or "barcode" (Aderem 2003). However, the idea of unique molecular patterns specific to pathogens which the immune system could recognize, so-called pathogen-associated molecular patterns (PAMPs), had to be relinquished, as the very same molecular patterns can be present and recognized in non-pathological contexts. Accordingly, they have been renamed "microbe-associated molecular patterns" (MAMPs) (Ausubel 2005; Koropatnick et al. 2004). The same molecular patterns can or will not induce an immune response in different contexts and the mapping between self/non-self and to what the immune system does or does not respond does not fully match.

An important though somewhat controversial modification was Polly Matzinger's "danger theory" (Matzinger 1994, 2002). Rather than reacting to anything foreign, the immune system would respond to anything 'dangerous'. But how to define 'danger'? And can it be linked to any specific molecular patterns? All these approaches try to link immunological recognition and response to molecular patterns associated with pathogens or danger. Applying metaphors and anthropomorphic concepts to describe certain molecules or microbes as being 'foreign', 'dangerous', or 'damage-associated' is

misleading as these concepts still fail to explain several phenomena.[69] To fill the explanatory gap, in their "discontinuity theory" Pradeu et al. suggest considering changes in the dynamics of such patterns (Pradeu et al. 2013; Pradeu and Vivier 2016). Greslehner explores the possibility that the immune system observes and responds to (microbial) functions rather than structures (Greslehner 2020). There is as yet no new uniformly accepted 'general theory of immunity' but theoretical progress continues to advance our understanding of how the immune system operates both on different levels and together with many other physiological systems (Eberl and Pradeu 2018).

At the same time, it is important to emphasize that the immune system executes many other functions in addition to host defense. The immune system is responsible for and involved with tissue repair, the clearance of damaged or dead cells and debris, developmental processes, the maintenance of homeostasis, and many more (Laurent et al. 2017; Pradeu 2019; Rankin and Artis 2018; Tauber 2017). Some of these immune functions are, in fact, carried out by non-immune cells, including microbes, thus leading to the "co-immunity" (Chiu et al. 2017) of a host together with its microbiota. In a similar vein, there has been a growth in the popularity of pursuing a systems biology approach (Davis 2020) and considering "the whole body as the system in systems immunology" (Poon and Farber 2020). Taken together, one should not consider 'the immune system' as a single defensive entity, but rather take seriously the 'system' in 'immune system'.

There have been a number of theoretical changes which have broadened our view of immunity and of the role of immune systems in health and disease in general (Tauber 2017). Many puzzles remain, but the last decade or so has seen enormous conceptual change in immunology. Thinking about the immune system in binary terms - like self/non-self, pathogenic/commensal - does not hold up anymore. These are fuzzy notions without clear-cut borders, often tied to metaphorical language.

In the following section we discuss this by highlighting three important features of the immune system, thereby presenting our conceptual model of the immune system.

5.3. Contextuality, regulation, and trade-offs

To better understand the nature of the immune system in general terms and to avoid misunderstandings, we propose thinking in terms of contextuality, regulation, and the trade-offs affecting immune responses.

5.3.1. Contextuality of the immune response

Rather than the immune system being activated only occasionally when facing threats, the immune system is in fact constantly interacting with its environment,[70] with the outcome of these interactions being context-dependent through and through. Furthermore, it is important to point out that such contextuality comes in many layers.

On the most general level, the contextuality of an immune response concerns the particular function at play, whether that be defense, tissue repair, the maintenance of homeostasis, the clearance of

[69] E.g. certain molecules, such as lipopolysaccharide, which can be found on the surface of both pathogens and commensals (Steimle et al. 2016); or flagellin, which may or may not trigger an immune response, depending on context (Park et al. 2019); and general puzzles concerning allergies and autoimmune diseases.
[70] This is perhaps most vividly illustrated by those mucosal surfaces which constantly interact with the microbiome but it also holds true for systemic immunity (Eberl 2016).

debris, including senescent cells, or a role in development. Which of these functions is triggered depends on the particular situation, driven by the integration of various signals and immune mediators such as cytokines.

One and the same thing can and often does fulfill different general functions, depending on what is going on. For instance, the triggering of Toll-like receptors (TLRs), a group of immune sensors specialized in recognizing microbial patterns, often leads to an inflammatory response under pathological conditions. However, TLR signaling is also crucial for maintaining intestinal epithelial homeostasis (Rakoff-Nahoum et al. 2004), a type of phenomenon called physiological inflammation (Sansonetti and Medzhitov 2009). Similarly, IFN-β, a type I interferon (IFN-I), is produced by epithelial cells and specialized subsets of immune cells early in a response to a viral infection. The infected cells start producing IFN-Is, the actions of which then interfere with viral replication in many ways. In addition to its presence in increased concentration during infection, IFN-β is also constitutively expressed at low levels and contributes to tissue homeostasis (Stefan et al. 2020).

Similarly, many immune cells enact various general functions, depending on the context. As with many other (if not all) immune cells, macrophages exhibit cellular plasticity, i.e., they change their phenotype depending on the particular context. The classically activated M1 macrophage phenotype leads to the triggering of an inflammatory response, whereas the alternatively activated M2 macrophage phenotype promotes tissue remodeling, among many other things (Biswas and Mantovani 2010).

Contextuality also pertains to the specifics of a given immune function. For instance, defense is often thought of in terms of resistance, i.e., the clearance of pathogens. However, an organism may instead prioritize another defense strategy called 'disease tolerance' (Medzhitov et al. 2012), a concept originating in the field of plant immunity. Whereas the resistance strategy is defined by reducing the pathogen burden, the consequence of which is always some degree of immunopathology, the tolerance strategy amounts to reducing the negative impact of pathogen-induced damage and immunopathology by decreasing the susceptibility of the host to tissue damage. Far from being a strict matter of an either-or strategy, resistance and tolerance may be located on a spectrum and, moreover, are pathogen-specific. For example, fatigue-induced anorexia, a kind of sickness behavior associated with infection, increases the tolerance to infection by *Salmonella typhimurium* while it decreases resistance to infection by *Listeria monocytogenes* in *Drosophila melanogaster* (Ayres and Schneider 2009). Morbidity and mortality in an infection may be due to a failure in resistance. However, if a comparable pathogen burden is found in hosts with different morbidity or mortality profiles despite the evidence of effective resistance, the pathology may result from a failure in tolerance (Medzhitov et al. 2012). Some studies have shown that there may be no significant difference in viral load in symptomatic versus asymptomatic cases of COVID-19 (Lee et al. 2020), meaning that the course of disease may, at least to some degree, reflect individual differences in susceptibility to tissue damage (Ayres 2020).[71]

Furthermore, although some microbes may exclusively be considered as pathogenic in humans, e.g., SARS-CoV-2, a large number of microbes are pathogenic only under certain conditions. In fact, pathogenicity is a complex and dynamic relation between the host and the microbe (Méthot and Alizon 2014). Only within the last few decades has the importance of the microbiome started to be fully appreciated in health and disease (Turnbaugh et al. 2007), offering additional "holobiont" or "superorganism" perspectives in immunology (Eberl 2010). Many viruses also exhibit interesting

[71] Similarly, Medzhitov et al. (2012) have proposed that the concept of tolerance may apply to phenomena such as the "Typhoid Mary", i.e., cases where a carrier remains healthy (asymptomatic) perhaps due to having a high level of tolerance to the particular pathogen.

contextual features, even though they have been predominantly associated with purely pathogenic or otherwise detrimental effects on the host. When we think of viruses, one immediately thinks of pathogens causing diseases which scourge humanity. However, viruses are also an oft-neglected, key part of the microbiome, ignored for a number of reasons which include not only methodological difficulties but also our biased perception owing to biomedical microbiology mostly being driven by a desire to understand pathogens. They not only play a role as pathogens; in the study of good health, the human virome is also a central factor which has until recently remained largely unexplored, remaining viral "dark matter" (Liang and Bushman 2021). Pradeu (2016) provides an intriguing overview, showing that while many viruses have become indispensable to host development, others confer protection against disease: for example, although many of the herpesviruses put individuals at risk of developing diseases, in their latent form several of the herpesviruses also provide protection against some bacterial infections such as by *Listeria monocytogenes* or *Yersinia pestis,* in an antigen non-specific way by upregulating the basal activation state of innate immunity (Barton et al. 2007). Furthermore, a virus may cause no clinical disease in a given species while in another it may induce severe disease. Because virulence is a function of traits that are intrinsic not only to the virus but also to the host, viruses that jump species may give rise to differences in clinical manifestation which most likely reflect the differences in host tolerance, provided that no evolution of the virus has occurred (Medzhitov et al. 2012). Owing to many of their features, including their innate immune system characteristic of the dampened activation of the inflammasome complex, bat species do not develop clinical disease despite harboring many viruses (Irving et al. 2021).[72] SARS-CoV-2 probably originated in bats, and although an evolutionary shift from its ancestor is likely, the resulting zoonosis is most likely due to the differences in the hosts' features including the mechanisms of disease tolerance.

The type of immune response to a microbe – such as type 1, type 2, or type 3 – is influenced by the lifestyle of the microbe. Roughly, type 1 immune response is aimed at intracellular pathogens, type 2 at extracellular parasites, and type 3 at extracellular bacteria and fungi. There are well-documented feedback mechanisms that ensure that the activation of one type of response inhibits the activation of another type of response, constantly balancing each other out, according to Eberl's equilibrium model of immunity (Eberl 2016). We shall return to this point when discussing the importance of regulation. It is not uncommon for the type of response directed against a microbe to change during the course of an infection.

Another 'contextuality layer' concerns the nature of the immune system itself. The immune system, far from being monolithic, consists of a vast network of interacting parts that can be carved up in multiple ways, most commonly into the humoral and cellular arms of the immune system, the barrier, innate and adaptive arms, or into the mucosal and systemic immune system.[73] As seen above, the

[72] Note that there are some differences between bat species. Furthermore, there are some rare examples of viruses that do cause severe disease in bat species.

[73] Although these categories do reflect the natural order of things to some extent, any such division is, however, always somewhat sketchy at best. For example, one of the defining features of the adaptive immune response is the development of immunological memory. However, it turns out that the innate arm also exhibits a memory phenotype, known as trained immunity (Netea et al. 2020). The humoral/cellular distinction is also fuzzy since in many cases there is no clear-cut difference between purely humoral and cell-mediated responses. Similarly, although the mucosal system can be characterized by a set of specialized lymphoid tissues, a specific circulating pattern, and a specific environment, the mucosal/systemic divide, according to which the influence of one upon the other may be severely limited, is also not as straightforward as sometimes believed. For instance, the phenomenon of oral tolerance pertains to the induction of a tolerogenic rather than an effector response toward harmless antigens at mucosal surfaces and can have systemic effects (Weiner 2000). Thus, many of these differences are, to some extent, a matter of degree rather than a matter of kind.

immune system can also be carved into different effector modules, i.e., types of responses, which are collections of humoral and cell-mediated mechanisms and span both the innate and adaptive arms.

The contextuality of the immune system and its environment has consequences for what attitude we ought to adopt toward terms such as 'strong', and 'weak', or 'weakened immune systems'. Casting our minds back to the diverse functions of the immune system, what would the concept of strong immunity mean in the context of homeostasis? Very little. Or consider the question of tissue repair. A 'strong immunity' in that particular context would probably amount to excessive tissue repair, giving rise to fibrosis (Medzhitov et al. 2012).

When thinking about the function of defense one often encounters the distinction between an immunocompetent and an immunocompromised individual. Providing a definition of an immunocompetent individual proves difficult and is hardly ever the subject of debate. Instead, an implicit reliance upon a *negative* definition – an individual who exhibits no (known) deficiency or is not immunosuppressed – seems to be the rule. However, the different outcomes of an immune response across individuals also result from various polymorphisms such as in the human leukocyte antigen (HLA) loci; these polymorphisms are not defects. Moreover, immunocompetence is also not something 'static' or 'given' because it evolves over time, most notably during development and aging. Intriguingly, the temporal changes in the workings of the innate and adaptive immune system, with consequences for its functions, also relate to circadian rhythms, i.e., they oscillate between day and night regimes (Druzd et al. 2017; Keller et al. 2009).

Yet immunocompromised individuals are characterized by the presence of some sort of immunodeficiency or by being in an immunosuppressed state. Since immunosuppression will be brought up in the next section, we shall leave discussion until then. Immunocompromised individuals have been intuitively considered as individuals with 'weak immunity'. However, closer inspection will reveal substantial problems with such an intuition.

With respect to immunodeficiency, it is customary to distinguish between primary or inherited immunodeficiency, and secondary or acquired immunodeficiency. An example of a primary immunodeficiency is a mutation in the gene encoding the transcription factor FoxP3 which plays a prominent role in the development and functioning of regulatory T cells. As a result, the suppressive function normally displayed by these cells is impaired and an autoimmune disease called IPEX (immune dysregulation, polyendocrinopathy, enteropathy, X-linked) develops (Rich et al. 2019). On the other hand, secondary immunodeficiencies, which are quite common, may arise from various causes. A secondary immunodeficiency may be due to an infection by, for example, HIV, which unless treated, normally leads to the acquired immunodeficiency syndrome (AIDS). Alternatively, it may be due to surgery. For example, a thymectomy or splenectomy, i.e., the removal of the thymus or the spleen, respectively, impairs some of the functions of the immune system. More specifically, a neonatal thymectomy prevents the development of mature T cells whereas a thymectomy in adulthood has little impact, since the pool of naïve T cells forms early on, and the thymus deteriorates with age, beginning soon after puberty. Perhaps less dramatically, the removal of the spleen confers a lifelong susceptibility to devastating infection by encapsulated bacteria such as *S. pnemoniae*, which requires that the affected individuals take antibiotics prophylactically and are vaccinated against pneumococcal infection. Individuals without a functional spleen lack the mononuclear phagocytes normally found within the spleen which clear this organism from the blood. However, such individuals are fully competent in launching a response against many other pathogens including many viruses, just like immunocompetent individuals.

Clearly, then, as the examples of neonatal thymectomy and splenectomy illustrate, it makes little sense to think in terms of there being a scale on which one may move between weakness and strength: it is simply either there, or missing. This is because the removal of these organs and the resulting deficiency in the development of particular cell populations or the impaired function of blood filtration, respectively, is not a 'thing' that can somehow be manipulated to be made 'stronger', since the thing to be made stronger is missing. Other times, it may be intuitive to think in terms of such a continuum between strong and weak, since, for instance, nutrition can be progressively improved in some sense. Malnutrition – another example of a secondary immunodeficiency – is known to affect cell-mediated immunity and is a major risk factor for many infectious diseases. Thus, intuitively we would believe that the better the nutrition the 'stronger' the immune protective effect. However, in contrast to this general belief, in certain specific circumstances deficiency may actually confer some additional level of protection. In particular, iron, though required in many immune-related pathways, happens to be essential for many bacteria, fungi and protozoa: it turns out that a certain degree of iron deficiency defined using ferritin and transferrin saturation in African children reduces the growth rate of the causative agent of malaria (Muriuki et al. 2019). Thus, the assumption that a good diet is always associated with increased protection turns out to be wrong in at least some, albeit very specific, cases. Still, this example illustrates yet again the crucial importance of contextual thinking.

There are at least two additional reasons to think that, in general, immunodeficiencies cannot be equated with an immune system being 'weaker' than that of an immunocompetent individual.[74] First of all, it is wrong to think that an immunodeficiency necessarily confers a system-wide defect. It is true that some defects (such as IPEX) prove fatal and others, such as a variety of severe combined immunodeficiencies (SCID), leave the host extremely susceptible to a wide range of conditions. However, other defects make the host overtly susceptible only to specific infections, while some may not even clinically manifest themselves. Some immunodeficiencies, such as the relatively common deficiency of IgA production, does not leave most of the individuals overly susceptible to infection, possibly owing to the compensation by IgM secretion (Yel 2010). Perhaps more strikingly, patients suffering from SCID were found to lack all innate lymphoid cell (ILC) subsets. After receiving hematopoietic stem cell transplantation and restoring T and B cell function, the ILC count was still considerably lower. However, the low count of ILCs was not associated with any particular susceptibility to disease even after decades of follow-up, suggesting that ILCs may exhibit some degree of redundancy (Vély et al. 2016).

The fact that a great many immunodeficiencies do not manifest themselves in a clinically relevant manner may be explained by the fact that the immune system, as with many other biological systems, is adaptive and notorious for exhibiting biological redundancy, meaning that, in many instances, should one pathway or one 'player' fail, another may step in and take over. Such redundancy thus gives rise to the phenomenon of robustness, that is, that the immune system is, on average, capable of functioning adequately even if some parts exhibit certain defects.[75]

[74] Note that this claim goes directly against public statements by distinguished institutions such as CDC (2021) which suggest that immunocompromised people possess a 'weakened' immune system. It is true that being immunocompromised is generally undesirable but the claim that defects somehow make one's immune system weaker is misguided, as we argue in the main text.

[75] What it means for a system to be robust, and what it means for it to exhibit redundancy, are complex questions since these concepts are applied to a wide range of phenomena. For instance, drawing on Kitano (2004), Truchetet and Pradeu (2018) define robustness as the maintenance of specific functionalities of a given system against internal and external perturbations. Truchetet and Pradeu also analyze robustness in pathological conditions, distinguishing robustness as dysfunction, when robustness is 'hijacked' in conditions

The second reason why one ought to avoid equating immunodeficiencies with 'weakness' is that although an immunodeficiency does indeed refer to a defect, that defect may result in an unwanted response that is 'too strong', i.e., one that is not kept in check. Recall the systemic autoimmune disease IPEX, a rare immunopathological condition in which the crucial suppressive function of regulatory T cells is impaired due to a mutated gene. Consequently, the affected individual suffers from a host of conditions including lymphoproliferation, thyroiditis, insulin-dependent diabetes mellitus, enteropathy, and other immune disorders (Rich et al. 2019). Thus, both primary and secondary immunodeficiencies cannot simply be understood as conditions related to the simplified notions of weakness or strength.

In summary, there are several important lessons to draw from a consideration of the contextual nature of the immune system. The immune system should not be seen as monolithic, i.e., one thing that performs in the same way across the whole spectrum of conditions. Rather, the immune system may better be considered in relative terms, in which the particular immune response depends upon a great many factors and changes accordingly. In one context, the immune response may display as effective and beneficial, whilst in another context, the response is inefficient and detrimental. To further see why it is so, we must consider the crucial role played by the adequate regulation of an immune response, to which we turn next.

5.3.2. Regulation of the immune response

During the course of an (inflammatory) immune response, the various mediators involved trigger a cascade of events leading to a build-up of molecules and cells in various tissues which further amplify the response by recruiting more and more immune mediators. As is the case with any such cascade, however, there must be a way to keep it from spiraling out of control. A vast array of feedback mechanisms serves that very purpose. In other words, any immune function must be finely tuned and tightly regulated by both internal and external signals.

Regulation takes place on multiple levels of organization. For instance, the complement system pathways are regulated by the presence of a set of proteases in the plasma or molecules constitutively expressed on cell surfaces. Co-stimulatory molecules provide an additional check on the activation of many types of immune cells by ensuring that an immune response is triggered in an appropriate context. Indeed, possibly unwanted responses are prevented by these mechanisms. The proper trafficking pattern is regulated by specific adhesion molecules and chemokines and their receptors. Cytokines are another major player in the regulatory processes as they influence cell responsiveness, proliferation, and differentiation. A specialized subset of regulatory T cells is necessary for the correct functioning of the immune system. In fact, recent studies have shown that many, if not all, types of immune cells exhibit both an effector and a regulatory phenotype (Alhabbab et al. 2019; Mantovani et al. 2011; Murray and Wynn 2011; Vivier et al. 2008).

Malfunction of the immune regulatory mechanisms – immune dysregulation – is at the heart of many pathologies. Rather than triggering an inflammatory response, in many cases the appropriate response is that of immunological tolerance, a concept that is different from the disease tolerance introduced above. It is easy to see how undesirable might be an inflammatory response against harmless food

such as AIDS and some types of cancer, and dysfunctional robustness, when a system should be robust but is not, in cases such as tissue repair.

metabolites: just think of food allergies.[76] In another case, pregnancy, a response directed against the embryo would result in miscarriage via the process of immunological rejection; and to cite but one more, a response against commensal microbiota leads to inflammation of the gut. In all these and numerous other cases, the immune response must be anti-inflammatory. Note that tolerance against 'non-self' is desirable here, not a malfunction. In yet other cases such as cancer, the induction of immunological tolerance by the tumor microenvironment is detrimental to the host.

The following example illustrates how regulation is also connected to the contextual nature of the type of an immune response. Experiments with mouse models of infection by the protozoan parasite *Leishmania major* showed what the adverse consequences of inappropriate cross-regulation may be. To clear the infection, a type 1 response, which includes the activation of T_H1 cells (which subsequently increase the activation level of macrophages) is required. Interestingly, while some mouse strains such as the C57BL/6 mice produce such a response, others such as the BALB/c mice trigger a type 2 response instead, with the differentiation of CD4 T cells into T_H2 cells which are unable to activate macrophages. Consequently, the mouse fails to clear the parasite and dies (Julia et al. 2000).

Several cues suggest that an improper cross-regulation – combined with certain kinetics of the responses – may also arise in some cases of COVID-19, where an increase in type 2 effectors has been observed in severe COVID-19, in contrast to a burst of type 1 and type 3 responses followed by their subsequent progressive reduction in moderate COVID-19 (Lucas et al. 2020).

While some dysregulation may be transient, e.g., due to a lack of those nutrients required primarily to maintain the function of a specific cell population (secondary immunodeficiency), dysregulation may also be persistent, such as on account of a mutation in a molecular regulator of inflammation which gives rise to autoinflammatory diseases characteristic of inflammation even in the absence of infection. Some forms of dysregulation may have recurrent clinical manifestation, whereas some others can remain hidden until showing in a particular clinical context. The latter has been found in COVID-19: some of the patients developing severe COVID-19 harbor antibodies against their own IFNs, in particular IFN-α2 and IFN-ω (Bastard et al. 2020). As discussed above, type I interferons are important early in an anti-viral response. However, the presence of autoantibodies against IFNs found predominantly in subsets of male patients with severe COVID-19 leads to the limited availability of IFNs and results in a delayed response and an improper recruitment of other immune cells. Combined with the fact that SARS-CoV-2, much like its predecessor SARS-CoV-1, appears to be a poor inducer of type I and type III IFNs responses (Blanco-Melo et al. 2020), possibly owing to the evasion strategy exhibited by SARS-CoV-2, using the papain-like protease SCoV2-PLpro (Shin et al. 2020), the autoantibodies against type I IFNs contribute to an imbalanced host response. Furthermore, a greater variety of other autoantibodies against immunomodulatory proteins have been found in patients with COVID-19 compared with uninfected controls, with the analysis suggesting the existence of both pre-existing and newly induced autoantibodies following the infection (E. Y. Wang et al. 2020).

Functional autoantibodies can be a sign of autoimmunity which may give rise to an autoimmune disease in which the immune system responds 'too strongly' against 'self' antigens, breaking the mechanisms of immunological tolerance. However, the outcome of such a response may result in a defective – perhaps 'too weak' – response in another context. This is what has been observed in COVID-19, but examples abound. For instance, autoantibodies targeting neutrophils cause neutropenia, i.e.,

[76] Note that food allergy is an immunological phenomenon, to be distinguished from food intolerance. While the former is an immune-mediated phenomenon, a type of a hyperresponsitivity, the latter pertains to a defect in metabolism, most commonly caused by a non-functional enzyme. Intolerance, in this sense, is to be associated neither with a failure of immunological tolerance nor a failure of disease tolerance.

the depletion of neutrophils, which leaves the individual particularly susceptible to infection by pyogenic bacteria; one of the therapeutic approaches is the removal of the spleen, which plays a major role in clearances.

Beyond the dysregulation caused by autoantibodies, there is another kind of dysregulation which appears to play a role in COVID-19 and beyond. It is now well established that the elderly exhibit low-grade chronic inflammation, which, although it does not appear to cause clinical problems, does contribute to disease. In general, the very young as well as the old are more susceptible to infections. However, this simple fact is not meaningfully captured by the concept of a 'weaker immunity'. Rather, the aging organism exhibits an immunosenescent phenotype of the innate immune system, characteristic of the condition of inflammaging which contributes to dysregulation by creating a constitutive pro-inflammatory environment (Shaw et al. 2010). As a result, some of the responses of the aging organism are improperly *enhanced* – hence dysregulated – which, together with other age-related changes such as the shift in the relative numbers of some immune cell subsets and their phenotypes may, at least in part, explain why age is a major risk factor in COVID-19 (Schultze and Aschenbrenner 2021).[77]

To see what an overly active immune response can accomplish, consider an extreme case of an inflammatory response going haywire – a *cytokine storm* – also heavily debated in the context of COVID-19. While some studies maintain that conceiving of COVID-19 as an instance of a cytokine storm may be a mischaracterization (Remy et al. 2020), others note that "no single definition of cytokine storm or the cytokine release syndrome is widely accepted, and there is disagreement about how these disorders differ from an appropriate inflammatory response" (Fajgenbaum and June 2020, p. 2255). One way or another, there appears to be agreement on one fundamental aspect of COVID-19: put eloquently, "the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has reminded us of the critical role of an effective host immune response and the devastating effect of immune dysregulation" (Fajgenbaum and June 2020, p. 2255).

The crucial regulatory aspect of the immune response suggests that a physiologically adequate response must be neither too strong, nor too weak, but just about right (Sakaguchi 2006). What 'right' means here depends not only on the proper regulation of the response but also on the specific kind of response which must be adjusted to the particular task, i.e., the contextual nature of what is going on. A response that is too vigorous results in much tissue damage. A response that is too permissive may not clear the pathogen, which may then establish a chronic infection, such as by SARS-CoV-2 (Kemp et al. 2021).

The importance of the regulatory processes can also be illustrated with reference to the maintenance of homeostasis, tissue repair, and disease tolerance. Mice with impaired TLR signaling cannot control homeostasis (or the development and maturation) of intestinal epithelium; such failure leads to chronic inflammation of the gut and the associated tissue damage as seen in inflammatory bowel diseases (Rakoff-Nahoum et al. 2004). Regulatory T cells have been shown to play an important role in promoting muscle repair and reducing inflammation upon injury in mice; depleting these cells leads to a disorganized tissue structure (Burzyn et al. 2013; Truchetet and Pradeu 2018). Similarly, just as a dysregulated immune response leads to pathology, disease tolerance must also be controlled if a

[77] Interestingly, evidence of the sustained immune dysregulation of several cell subsets has been found in both hospitalized and non-hospitalized infected individuals. Moreover, some of these changes were found to increase in time in non-hospitalized patients. The lasting effects on subsequent infections or inflammatory diseases are yet to be determined (Files et al. 2021).

pathology is to be avoided, e.g., fibrosis resulting from an excessive – dysregulated – tissue repair (Medzhitov et al. 2012).

Although the immune system appears sophisticated and effective, it nevertheless does not always get things 'right' (from our human perspective), needing to be nudged in the right direction. Drugs with immunomodulatory effects serve precisely this purpose. In cases in which an immune response is 'too weak', it is possible to artificially *stimulate* or *boost* it, i.e., not just make it 'stronger' but to specifically intervene and trigger it or facilitate certain responses. For instance, cancer patients may now benefit from a variety of immunotherapies, including immune checkpoint inhibitors (Ribas and Wolchok 2018). Tumors often induce an immunosuppressive environment through a number of different mechanisms, including making use of normal regulatory mechanisms such as the expression of the CTLA-4 molecule on activated tumor-specific T cells. This molecule can be targeted by the drug ipilimumab, allowing these T cells to overcome their normal physiological limits and contribute to tumor elimination. However, such manipulation of the immune system comes at costly side-effects that must be addressed, such as the onset of autoimmune diseases in a subset of patients. The immune system also often fails to respond against purified proteins. Rather, because such proteins are usually poorly immunogenic they often induce a state of immunological tolerance. In many contexts this is usually beneficial. However, it presents an obstacle in the design of vaccines based on the use of purified proteins, including vaccines based on the toxoid design like the tetanus toxoid, or the subunit vaccines. To overcome these difficulties, scientists have developed a number of adjuvants – substances that increase reactivity, most notably by stimulating innate sensor pathways. Although many adjuvants are routinely used in experimental research, only a few are approved for clinical use in humans. The problem is that most adjuvants cause dangerously excessive inflammation. Thus, boosting an immune response must always be kept within strict limits.

Consider also the case of disease tolerance. As Medzhitov et al. (2012) argue, boosting an immune response when the problem is a failure of tolerance may prove ineffective or even detrimental, whereas boosting tolerance may provide health benefits by limiting tissue damage caused either by the pathogen directly or by the immune response to the microbe. However, here again it is important to stress that any such action must be carefully regulated. Although some tolerance mechanisms appear to be at work at the basal level, others are inducible and work at the expense of normal tissue function. Thus, much like the mechanisms of resistance, tolerance also comes at a cost. Furthermore, tolerance mechanisms also require tight control; otherwise, they result in pathology, as illustrated by the above example of fibrosis.

In other cases, it is desirable to *attenuate* an immune response rather than boost it. Transplant patients take immunosuppressive drugs in order for the transplanted organ not be rejected. Dexamethasone, a synthetic chemical similar to cortisol with potent immunosuppressive effects, has found its use in the treatment of some autoimmune diseases and it has also shown clinical benefit in patients with COVID-19 who require oxygen therapy or mechanical ventilation (The RECOVERY Collaborative Group 2021). However, to see how complex things get, consider again the case of IFNs. As we saw, part of the problem in some severe cases of COVID-19 may be a delayed anti-viral response due to the presence of autoantibodies against subsets of IFNs (e.g., IFN-α2). Since administering exogenous IFN-α is considered unlikely to provide benefit (Bastard et al. 2020), some are testing a treatment with IFN-β (Bastard et al. 2021), the motivation being that IFNs may enhance the immune response in non-specific ways. In fact, IFN-β is used in the treatment of diseases of viral origin (Guarda et al. 2011). However, IFN-β has also been found to reduce a response rather than to enhance it, and as such it has proven useful in treating patients with multiple sclerosis, owing to its effect on the reduction of IL-1 production, thereby limiting a powerful mediator of inflammation (Guarda et al. 2011).

Chemotherapy provides another interesting example. While it is true that chemotherapy is generally associated with an increased risk of infection due to its immunosuppressive effect and as such it is often categorized as a secondary immunodeficiency, it has also been found that certain chemotherapeutic drugs such as anthracyclines work, in part, by increasing the immunogenicity of the tumor cells, thus increasing the anti-tumor responses (Alizadeh and Larmonier 2014). Consequently, chemotherapy may be said to result in general suppression by lowering the count of immune cells, while at the same time it may boost a specific immune response, provided that enough antigen-specific T cells survive the therapy.

It should also be noted that the idea of modulation presupposes that there is something to be modulated. In the above cases, this assumption was implicit. Given that the immune system is not monolithic, but rather consists of a large number of diverse molecules, cell populations, and several kinds of specialized organs, it may not always be the case that there is something to be modulated. Indeed, consider again the removal of the thymus at an early age which prevents the development of mature T cells: there is nothing left to modulate.

5.3.3. Trade-offs

Given the complexity of the biological systems and the discussion in the two previous sections, it should come as no surprise that the workings of the immune system exhibit numerous trade-offs on multiple levels of organization.

There are trade-offs between a beneficial function under some conditions and a detrimental effect under other conditions. For instance, recall the debate on cell plasticity: the M2 phenotype of macrophages is crucial in the process of tissue repair but its presence in cancer is generally associated with pro-tumoral effects (Biswas and Mantovani 2010; see also Truchetet and Pradeu 2018). Similarly, following liver damage, the transient induction and accumulation of senescent cells help to resolve fibrosis (Krizhanovsky et al. 2008). However, senescent cells need to be cleared by the immune system since their prolonged existence is considered detrimental. In the aging organism such detrimental effects become apparent as these cells accumulate, owing either to a decrease in the clearance capabilities of the immune system or to an increase in the generation of such cells which exceeds the capacity of the immune system to clear them (Rodier and Campisi 2011).

Turning to the function of defense, it is important to realize that there is a trade-off between resisting an infection, i.e., the clearance of pathogens, and the tissue damage arising from the immune response, i.e., the immunopathology. A ‘strong’ response, in this sense, may be associated with the vigorous clearance of pathogens while giving rise to a cytokine storm, an immune-mediated life-threatening condition.[78] Yet another important kind of trade-off is made with respect to evolutionary fitness. The host has to find a balanced immune response and allocate the resources and energy, as any immune response comes at a cost (Lochmiller and Deerenberg 2000). On the flip-side of this coin we find the trade-offs with respect to parasite virulence (Alizon et al. 2009).

Some trade-offs can also be viewed in terms of something being incompatible with something else. For instance, Medzhitov et al. (2012) note that a response against microbe A can be incompatible with tolerating microbe B, giving rise to the phenomenon of negative preconditioning. They also note that

[78] Some authors argue that the trade-off between resistance and immunopathology can be resolved, to some extent, by tolerance mechanisms. Since these limit tissue damage, they allow for a prolonged duration of the immune response (Medzhitov et al. 2012).

coinfection, e.g., a viral infection of the respiratory tract followed by a respiratory bacterial infection, often results in severe morbidity and mortality which is usually thought to be the consequence of compromised immunity. Indeed, Eberl (2016) proposes that the cross-regulation of types of responses may be the problem, i.e., viral infection induces a type 1 response which inhibits the type 3 responses crucial for clearing extracellular bacterial infections. However, as Medzhitov et al. suggest, it is also possible that the inducible tolerance to the particular viral infection is incompatible with tolerance to the respiratory bacterial infection which may be why that kind of coinfection is dangerous.

Similarly, while some disease tolerance mechanisms are constitutive, others are inducible and come at a cost: they work at the expense of normal tissue function. Thus, there is a trade-off of the incompatibility kind between normal tissue function and an increased tolerance to tissue damage.

Various trade-offs also arise when therapeutically manipulating the immune responses. Recall the use of ipilimumab in cancer treatment and its known side effect - the onset of severe autoimmunity in some cases. In order to avoid transplant rejection, patients receiving transplants are put on non-specific immunosuppressive drugs for life, which, however, leave them more susceptible to infection and cancer.

Finally, the idea that one's particular genotype can influence the susceptibility of an individual is also well established. However, the same genes that confer protection can also make the individual susceptible to other conditions. For example, using genome-wide association studies, the genetic variation in human leukocyte antigen (HLA) molecules has been established as one of the strongest predictors of HIV-1 control (Pereyra et al. 2010). While the HLA-B*27 allele, by virtue of its mechanistic function, has been found to increase resistance to HIV-1, it also leaves the host at greatly increased risk of developing ankylosing spondylitis, an autoimmune disease (Murphy and Weaver 2017). Trade-offs like these can be found anywhere you look.

5.4. The uses and abuses of metaphors

Metaphors abound in immunology, from prominent war metaphors of defense driven by the discovery of microbial pathogens, through the idea of an immunological 'self', to slogans about 'strengthening' or 'boosting' one's immune system. The use of metaphors has been deeply rooted in immunology throughout its history, not just as communication devices with a wider audience, but shaping the very way scientists think, understand, and build theories in immunology (Institute of Medicine 2006; Löwy 1996). In particular, theoretical metaphors of an immunological 'self' have become prominent and criticized (Pradeu 2012; Tauber 1994). The claim that science is soaked in metaphorical language is scarcely contested. Rather, "in science, metaphor is widely considered an essential tool for understanding" (Ball 2011). Metaphorical language is often used to understand and communicate complex phenomena that are not completely understood by referring to other, more familiar concepts: "*The essence of metaphor is understanding and experiencing one kind of thing in terms of another*" (Lakoff and Johnson 1980, p. 5, original italics).

It has been proposed that, in science, metaphors serve at least three functions which are often interrelated in various ways (Bradie 1999; see also Kampourakis 2020). Metaphors have a *heuristic function* which helps scientists explore new phenomena by referring to other, already understood phenomena. Such a function is also achieved by drawing on a variety of analogies, a practice which has been documented in empirical studies of how immunologists reason (Dunbar 2002).

Metaphors also have an indispensable *theoretical function*, i.e., they facilitate the understanding and explanation of phenomena. For instance, the self/non-self framework was put forth as an explanation of the basis of immune response. Notwithstanding their undisputable usefulness, it has also been well recognized that metaphors can also obscure understanding and lead one astray. As Philip Ball has put it:

> "Books of life, junk DNA, DNA barcodes: all these images can and have distorted the picture, not least because scientists themselves sometimes forget that they are metaphors. And when the science moves on — when we discover that the genome is nothing like a book or blueprint — the metaphors tend, nonetheless, to stick. The more vivid the image, the more dangerously seductive and resistant to change it is" (Ball 2011).

As we sought to show in Section 5.2., the self/non-self framework and the idea of the immune system as a defense system may be a victim of this kind of seductiveness.

Finally, metaphors have a *rhetorical function*: they play an important role in science education and communication. More importantly, rather than being merely instruments of getting a point across to a larger audience, "metaphors have profound influences on how we conceptualize and act with respect to important societal issues" (Thibodeau and Boroditsky 2011).

Putting all this together, let us assess the idea that there is such a thing as a strong or weak immune system or response. As we have argued, the immune system is not monolithic but rather a complex system composed of many parts, and involved in many functions, the outcome of which is context-dependent and exhibits trade-offs.

There does seem to be a perfectly legitimate *descriptive* sense in speaking of immune response when defined as pertaining to a quantitative measurement of certain immune features: e.g., assays allowing measurements of cytokine production, the number of cells, the titers of neutralizing antibodies as a proxy for protective immunity, binding affinities, and so on. Overall, however, the talk of a strong/weak immune system or response mischaracterizes the workings of the immune system. The following list derives from the discussion in previous sections and is in no way exhaustive.

(i) *Normative connotation*. Strong defense or the idea of boosting immunity may be viewed as desirable, but in many cases it may lead to pathology or come at a cost (think of a cytokine storm or checkpoint inhibitors).

(ii) *Paradoxical connotation*. Immunodeficiency invites the intuition that the issue is one of a weak response. However, the same immunodeficiency could also concern an issue of a strong response (think of IPEX) and it is not clear which notion should apply.

(iii) *Not applicable because not amendable to change*. Thinking in terms of a continuum between strong and weak immunity and the idea of boosting immunity is sometimes invalid (think of neonatal thymectomy).

(iv) *Lack of meaning conveyed*. Many phenomena and functions of immunity cannot be meaningfully captured by these notions (think of homeostasis).

(v) *Failure to account for what the immune system is, what it does and why it sometimes fails*. Rather than being in a steady state until an occasional threat emerges, the immune system is constantly active in maintaining various functions, including functions other than defense. All immune-related phenomena require a contextual understanding; otherwise, one would fail to understand why a phenomenon may be desirable but also detrimental (think of immunological tolerance or the trade-off between immunity and immunopathology). The notions of strong/weak immunity also give the false impression that the immune system can be described along this (one) dimension.

While stressing that the immune system is more than a defense system, defense remains one of its functions, and an important one at that. In immunology, however, defense and war metaphors have biased our perspective exclusively toward defense at the expense of paying attention to the numerous other roles that the immune system plays. Similarly, thinking in terms of a strong or weak immune system involves all kinds of misleading intuitions, as shown above. These intuitions have been perpetuated in certain areas of research, in science education (especially textbooks), and in public communication. One of the most difficult aspects of the COVID-19 pandemic has been the communication of science as well as the gaps in our understanding of the immune system (Nature Reviews Immunology Editorial 2021). Metaphors can guide our intuitions, but then "immunology is where intuition goes to die" (Young 2020). This clearly impacts how we estimate risks, behave, and what interventions we consider in attempting to improve our health - including the idea of "boosting the immune system" (Cassa Macedo et al. 2019).

Moreover, this line of thought can also be dangerous, such as when it provides us with a false sense of security. Of particular danger is the message that healthy people with a 'strong' immune system who lack pre-existing conditions are not affected. Equally, the idea of a 'weak' immune system is also mistaken. Therefore, being part of a risk group (e.g., due to some forms of compromised immunity to a disease such as COVID-19) generally does not depend on the 'weakness' of one's immune system. Also, things like proper vitamin intake or reduced stress levels are generally beneficial, but for reasons other than the weak/strong dichotomy. The popular idea that vitamins work by boosting one's immunity is false: for instance, vitamin D is thought to work by helping to balance an immune response, inhibiting certain responses and providing an anti-inflammatory response (Jain et al. 2020). Similarly, stress – often perceived as weakening the immune system – is rather an array of complex phenomena with multiple antagonistic effects (Dhabhar 2014).

In summary, we argue that the notions of strong and weak immunity would best be dropped and replaced by a more nuanced view based on the model of the immune system which takes into account the features of contextuality, regulation and trade-offs. Furthermore, the notions of strength/weakness invite too narrow a focus on defense as the only role of the immune system, at the expense of neglecting many other, equally vital immune functions. Finally, these notions miss out on the crucial importance of the immune system in maintaining balance and thereby preventing pathology.

5.5. Chapter summary

Several areas of contemporary research in immunology show that the immune system does much more than defend the host (i.e., self) against non-self. Metaphors are crucial to scientific endeavor and communication but they also sometimes mislead. Thinking in terms of defense invites notions such as strong or weak defense, while not adequately accounting for what the immune system is, what it does, and why it sometimes fails. Although there is a limited sense in which the notion of strong/weak or boosted/attenuated immunity makes perfect sense, in general it cannot account for the nature of most immunological phenomena. Instead, we propose a conceptual model, focusing on contextuality, regulation, and trade-offs, which give due credit to the complexity of the immune system. A better understanding of the immune system will allow us both to address open questions concerning COVID-19 and its long-lasting effects, and to prepare us for future pandemics.

Philosophy of science can help us here by clarifying immunological concepts, assessing assumptions and methods, formulating new concepts, models or theories (some of which suggest new experiments

or therapeutic targets), and by opening a dialogue not only between the sciences but also between the sciences and society (Laplane, Mantovani, et al. 2019).

Concluding remarks and future challenges

Scientific modeling is an important tool in contemporary science including biomedical research. A better understanding of modeling is therefore of crucial importance to any theory of science. In this PhD thesis I have provided an original contribution to several debates concerning the nature and use of scientific modeling.

I have proposed a novel account of scientific modeling – the *experimentation-driven modeling account* (EDM) – according to which scientists integrate piecemeal experimental results into a unified conceptual framework that is expressed in the form of a *mechanistic model*, most often in the form of a diagram. Although I outlined the EDM account against the backdrop of the research project focused on studying the role of myeloid-derived suppressor cells in the formation of the pre-metastatic niche, the account applies generally to such laboratory research which aims to construct mechanistic models of biological phenomena. As argued, there are epistemic similarities as well as important differences between the EDM account and what I call the *description-driven modeling account* (DDM). As we have seen, scientists involved in EDM are neither sociologically nor professionally recognized as modelers despite the fact that they propose various mechanistic models. Instead, the sociological dimension of modeling appears to exclusively apply to DDM. Therefore, the analysis provided in Chapter 1 may prove insightful when it comes to building interdisciplinary bridges between research approaches that are usually considered to be more formal and those that characterize wet labs. In other words, one could argue that by illuminating the two modeling accounts, such analysis may facilitate a mutual understanding between researchers trained in either tradition. Often, researchers involved in interdisciplinary groups – still rather rare – draw on an implicit understanding of the respective approaches; something that has been made more explicit herein. Thus, even those involved in interdisciplinary groups may benefit from such analysis. However, more work remains to be done. For instance, the future challenge pertains to providing an explicit consideration of the precise way in which experimental results are integrated in the process of model construction.

I have also discussed the issue of mechanistic models which provide mechanistic explanations of biological phenomena such as the models in molecular biology accounting for gene expression. In particular, I addressed an objection due to Alan Love and Marco Nathan raised against the mechanistic account of explanation according to which the account fails to accommodate the common practice of idealizing difference-making factors. I scrutinized the arguments and examples provided by Love and Nathan and I presented reasons for thinking that their analysis fails to provide support to their conclusion since their analysis paints a confusing picture as it is interwoven with conceptual inconsistencies. In Chapter 2 I predominantly focused on showing that one of the reasons for a thorough conceptual clarification is to avoid using arguments in philosophical discussions which are grounded in concepts that are ill-defined for the purposes at hand; failing to clarify such key concepts may generate great misunderstandings and lead to cycles of fruitless debates, generating even more confusion. Another reason for the endeavor of clarifying concepts concerns a more direct contribution to some of the methodological debates within scientific communities such as those over the issue of the *realisticness* of assumptions in providing understanding of phenomena; a problem also reflected in Chapter 4 regarding the use of assumptions in agent-based modeling of an epidemic. Throughout Chapter 2 I raised concerns about the development and subsequent use of the notions of abstraction and idealization in the context of a particular field and their direct application to a different field. This is because the respective fields may differ from each other to the extent that they exhibit significantly different epistemic practices. Indeed, I illustrated this point using the example of developing an account of abstraction and idealization in the context of certain kinds of models in evolutionary biology, which, arguably, are very different from the kinds of models used in molecular biology. In part,

this brings us back to the discussion in Chapter 1, namely the two approaches to modeling, with evolutionary models and molecular biology models exemplifying the DDM and EDM style, respectively. Thus, in the future work it may be worth to further explore different epistemic practices and thus shed some new light on the debate on abstraction and idealization, thereby illuminating the methodological debates within sciences over and above the purely philosophical ones.

When introducing the EDM account in Chapter 1 in order to provide an account of mechanistic model building in laboratory biomedical research I briefly discussed the use of animal models (mouse models in particular) and cell cultures. However, a more elaborate discussion of these experimental systems had to wait until Chapter 3 in which I analyzed the kinds of similarity judgments considered by cancer biologists and cancer immunologists. I argued that when choosing between existing mouse models, a variety of criteria ranging from the research question at hand and pragmatic and other factors in addition to similarity judgments enter the decision-making process. When extrapolating results obtained from a mouse model to a human, similarity judgments concern the evaluative aspect, i.e., the predictive accuracy of the mouse model, and thus the justification of extrapolative inferences. Finally, when creating a new mouse model, similarities pertain to the intention to actively introduce changes into mouse cohorts so that relevant similarities arise. The discussion in Chapter 3 helps to shed some light on the debate on similarity in scientific representation. In particular, I argued that whereas in model extrapolation the role of similarity suggests that a conceptual distinction between representation and accurate representation can be maintained by construing the latter in terms of predictive accuracy, it holds for neither model selection nor model creation. This is because in the two latter research modes, similarity considerations play a key role in the process of establishing and maintaining a representational relation.

While Chapter 3 expanded on EDM, Chapter 4 did the same with regard to DDM, drawing on joint work with Mariusz Maziarz. The COVID-19 pandemic has put a spotlight on the methodology of formal modeling and the use of assumptions, somewhat reminiscent of the discussion in Chapter 2. Rather than focusing on the most commonly used compartmental models, Chapter 4 analyzed agent-based modeling (ABMs). With the increase in computational power over the past several decades, ABMs present a relatively novel way to simulate the behavior of complex systems. ABMs consist of entities – namely agents – that interact both with each other and their environment according to a defined set of rules. Using assumptions and empirical data as inputs, scientists construct various ABMs to study the behavior of the model system. Thus, ABMs are an instance of the DDM practice. I discussed an example of an ABM, the SARS-CoV-2 version of the ACEMod, and I noted that such ABMs allow us to compare the baseline scenario (simulation of the epidemic with no changes to agents' behavior) with effects produced by alternative suppression measures. In so doing, ABMs allow us to assess counterfactual causal claims, thereby providing evidence for policy making. As any other model, ACEMod also relies on various simplifying assumptions. Notwithstanding these simplifications, I argued that the process of model calibration and validation suggests that the model in question represents the actual mechanism of the SARS-CoV-2 epidemic. Thus, provided that the model assumptions are calibrated and checked against the background empirical data - that is, the components, their activities, and spatiotemporal organization resemble (in relevant aspects and to a certain degree) the actual state of things - iterative runs of the simulations can indeed provide understanding and inform policy decisions. That being said, I also noted that further research is needed to assess the risk of bias in the epidemiological ABMs that deliver both difference-making and mechanistic evidence. This is because there still is a risk that the assumptions employed in the ABM in question do not include all the key factors, and that they make model predictions susceptible to the problem of confounding. Moreover, in reaching policy decisions, ABMs should be understood as merely one piece of the puzzle subject to further re-evaluation with respect to value judgments. This is because alternative mitigation

measures may disproportionately affect certain social groups. In light of the pandemic which has affected the life of almost everyone on the planet and which has caused great suffering, assessing the modeling practice while drawing on philosophy is an important endeavor.

As noted in the introduction, scientific literature is notoriously loose when it comes to providing a precise clarification of some of the general concepts such as 'model' or 'modeling'. The aim of Chapter 1 was to provide such clarification. Yet in Chapter 5 I used the term 'model' loosely to refer to the conceptual framework regarding the understanding of the immune system that I proposed based on a joint work with Gregor Greslehner. However, this is in line with the usual goal of a scientific paper which usually consists in addressing a scientific problem rather than the meaning of a general term that is largely unrelated to the issue at hand. The analysis provided in Chapter 5 should have demonstrated several things. Firstly, thinking in terms of defense invites notions such as strong or weak defense, neither of which adequately accounts for what the immune system is, what it does, and why it sometimes fails. Although there is a limited sense in which the notion of strong/weak or boosted/attenuated immunity makes perfect sense, in general it cannot account for the nature of most immunological phenomena. Instead, such thinking involves all kinds of misleading intuitions. Secondly, the proposed conceptual model that is intended to replace the misguided thinking discussed above turns the spotlight onto three crucial features of the immune response: contextuality, regulation, and trade-offs. A better understanding of the immune system will allow us both to address open questions concerning COVID-19 and its long-lasting effects, and to prepare us for future pandemics. Moreover, philosophy of science can help us here by clarifying immunological concepts, assessing assumptions and methods, formulating new concepts, models or theories (some of which suggest new experiments or therapeutic targets), and by opening a dialogue not only between the sciences but also between the sciences and society. Given the ever-increasing importance of science communication, this may possibly prove to be an important contribution to a burning issue surrounding the society since inadequate understanding often leads to inadequate behavior responses which affects both the individual engaged in such behavior and the society at large.

In summary, the topic of scientific modeling continues to be a fruitful object of philosophical reflection, with many issues still waiting to be resolved. In this PhD thesis I have taken it upon myself to shed new light on some of the most pressing issues including those that have a broader impact on society.

Bibliography

Acuna-Zegarra, M. A., Comas-Garcia, A., Hernandez-Vargas, E., Santana-Cibrian, M., & Velasco-Hernandez, J. X. (2020). The SARS-CoV-2 epidemic outbreak: a review of plausible scenarios of containment and mitigation for Mexico. *medRxiv*, 2020.03.28.20046276. doi:10.1101/2020.03.28.20046276

Adam, D. (2020). Special report: The simulations driving the world's response to COVID-19. *Nature*, *580*(7803), 316–318. doi:10.1038/d41586-020-01003-6

Aderem, A. (2003). Phagocytosis and the Inflammatory Response. *The Journal of Infectious Diseases*, *187*(s2), S340–S345. doi:10.1086/374747

Adlhoch, C., Baka, A., Cenciarelli, O., Penttinen, P., Palm, D., Plachouras, D., et al. (2020). *Contact tracing: Public health management of persons, including healthcare workers, having had contact with COVID-19 cases in the European Union.* https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-public-health-management-contact-novel-coronavirus-cases-EU.pdf

Alhabbab, R. Y., Nova-Lamperti, E., Aravena, O., Burton, H. M., Lechler, R. I., Dorling, A., & Lombardi, G. (2019). Regulatory B cells: Development, phenotypes, functions, and role in transplantation. *Immunological Reviews*, *292*(1), 164–179. doi:10.1111/imr.12800

Alizadeh, D., & Larmonier, N. (2014). Chemotherapeutic targeting of cancer-induced immunosuppressive cells. *Cancer Research*, *74*(10), 2663–2668. doi:10.1158/0008-5472.CAN-14-0301

Alizon, S., Hurford, A., Mideo, N., & Van Baalen, M. (2009). Virulence evolution and the trade-off hypothesis: History, current state of affairs and the future. *Journal of Evolutionary Biology*, *22*(2), 245–259. doi:10.1111/j.1420-9101.2008.01658.x

Andersen, H. (2012). Mechanisms: what are they evidence for in evidence-based medicine? *Journal of Evaluation in Clinical Practice*, *18*(5), 992–999. doi:10.1111/j.1365-2753.2012.01906.x

Andersen, H. (2014a). A Field Guide to Mechanisms: Part I. *Philosophy Compass*, *9*(4), 274–283. doi:10.1111/phc3.12119

Andersen, H. (2014b). A Field Guide to Mechanisms: Part II. *Philosophy Compass*, *9*(4), 284–293. doi:10.1111/phc3.12118

Anjum, R. L., Copeland, S., & Rocca, E. (2020). Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence.' *BMJ Evidence-Based Medicine*, *25*(1), 6–8. doi:10.1136/BMJEBM-2018-111092

Ankeny, R. A., & Leonelli, S. (2011). What's so special about model organisms? *Studies in History and Philosophy of Science Part A*, *42*(2), 313–323. doi:10.1016/J.SHPSA.2010.11.039

Ankeny, R. A., & Leonelli, S. (2020). *Model Organisms*. Cambridge: Cambridge University Press.

Ankeny, R. A., Leonelli, S., Nelson, N. C., & Ramsden, E. (2014). Making Organisms Model Human Behavior: Situated Models in North-American Alcohol Research, since 1950. *Science in Context*, *27*(03), 485–509. doi:10.1017/S0269889714000155

Atanasova, N. (2015). Validating Animal Models. *THEORIA. An International Journal for Theory, History and Foundations of Science*, *30*(2), 163–181. doi:10.1387/theoria.12761

Ausubel, F. M. (2005). Are innate immune signaling pathways in plants and animals conserved? *Nature Immunology*, *6*(10), 973–979. doi:10.1038/ni1253

Ayres, J. S. (2020). Surviving COVID-19: A disease tolerance perspective. *Science Advances*, *6*(18), eabc1518. doi:10.1126/sciadv.abc1518

Ayres, J. S., & Schneider, D. S. (2009). The Role of Anorexia in Resistance and Tolerance to Infections in Drosophila. *PLoS Biology*, *7*(7), e1000150. doi:10.1371/journal.pbio.1000150

Baetu, T. M. (2014). Models and the mosaic of scientific knowledge. The case of immunology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *45*, 49–56. doi:10.1016/J.SHPSC.2013.11.003

Baetu, T. M. (2016). The 'Big Picture': The Problem of Extrapolation in Basic Research. *The British Journal for the Philosophy of Science*, *67*(4), 941–964. doi:10.1093/bjps/axv018

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D.-Y., Chen, L., & Wang, M. (2020). Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA*, *323*(14), 1406. doi:10.1001/jama.2020.2565

Bailer-Jones, D. (2009). *Scientific models in philosophy of science*. Pittsburgh: University of Pittsburgh Press.

Ball, P. (2011). A metaphor too far. *Nature*. doi:10.1038/news.2011.115

Bankes, S. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*, *41*(3), 435–449. doi:10.1287/opre.41.3.435

Barton, E. S., White, D. W., Cathelyn, J. S., Brett-McClellan, K. A., Engle, M., Diamond, M. S., et al. (2007). Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature*, *447*(7142), 326–329. doi:10.1038/nature05762

Bastard, P., Lévy, R., Henriquez, S., Bodemer, C., Szwebel, T.-A., & Casanova, J.-L. (2021). Interferon-β Therapy in a Patient with Incontinentia Pigmenti and Autoantibodies against Type I IFNs Infected with SARS-CoV-2. *Journal of clinical immunology*, 1–3. doi:10.1007/s10875-021-01023-5

Bastard, P., Rosen, L. B., Zhang, Q., Michailidis, E., Hoffmann, H. H., Zhang, Y., et al. (2020). Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science*, *370*(6515), eabd4585. doi:10.1126/science.abd4585

Batterman, R. W. (2002). Asymptotics and the Role of Minimal Models. *The British Journal for the Philosophy of Science*, *53*(1), 21–38. doi:10.1093/bjps/53.1.21

Batterman, R. W. (2009). Idealization and modeling. *Synthese*, *169*(3), 427–446. doi:10.1007/s11229-008-9436-1

Batterman, R. W., & Rice, C. C. (2014). Minimal Model Explanations. *Philosophy of Science*, *81*(3), 349–376. doi:10.1086/676677

Bertolaso, M. (2016). *Philosophy of Cancer: A Dynamic and Relational View*. Dordrecht: Springer.

Biswas, S. K., & Mantovani, A. (2010). Macrophage plasticity and interaction with lymphocyte subsets: Cancer as a paradigm. *Nature Immunology*, *11*(10), 889–896. doi:10.1038/ni.1937

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Møller, R., et al. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, *181*(5), 1036-1045.e9. doi:10.1016/j.cell.2020.04.026

Bodenstein, M., Corsetti, G., & Guerrieri, L. (2020). Social Distancing and Supply Disruptions in a Pandemic. *Finance and Economics Discussion Series*, *2020*(031). doi:10.17016/feds.2020.031

Boesch, B. (2017). There Is a Special Problem of Scientific Representation. *Philosophy of Science*,

84(5), 970–981. doi:10.1086/693989

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, *97*(3), 303. doi:10.2307/2185445

Bokulich, A. (2011). How scientific models can explain. *Synthese*, *180*(1), 33–45. doi:10.1007/s11229-009-9565-1

Bokulich, A. (2017). Models and Explanation. In L. Magnani & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 103–118). Dordrecht: Springer. doi:10.1007/978-3-319-30526-4_4

Bolinska, A. (2013). Epistemic representation, informativeness and the aim of faithful representation. *Synthese*, *190*(2), 219–234. doi:10.1007/s11229-012-0143-6

Bolinska, A. (2016). Successful visual epistemic representation. *Studies in History and Philosophy of Science Part A*, *56*, 153–160. doi:10.1016/J.SHPSA.2015.09.005

Bolker, J. A. (2009). Exemplary and Surrogate Models: Two Modes of Representation in Biology. *Perspectives in Biology and Medicine*, *52*(4), 485–499. doi:10.1353/pbm.0.0125

Boone, W., & Piccinini, G. (2016). Mechanistic Abstraction. *Philosophy of Science*, *83*(5), 686–697. doi:10.1086/687855

Bradie, M. (1999). Science and Metaphor. *Biology & Philosophy*, *14*(2), 159–166. doi:10.1023/A:1006601214943

Broadbent, A., & Smart, B. (2020). Why a one-size-fits-all approach to COVID-19 could have lethal consequences. *The Conversation*. https://theconversation.com/why-a-one-size-fits-all-approach-to-covid-19-could-have-lethal-consequences-134252

Budhu, S., Wolchok, J., & Merghoub, T. (2014). The importance of animal models in tumor immunity and immunotherapy. *Current Opinion in Genetics & Development*, *24*, 46–51. doi:10.1016/J.GDE.2013.11.008

Burnet, F. M. (1969). *Cellular Immunology: Self and not-self*. Cambridge: Cambridge University Press.

Burzyn, D., Kuswanto, W., Kolodin, D., Shadrach, J. L., Cerletti, M., Jang, Y., et al. (2013). A special population of regulatory T cells potentiates muscle repair. *Cell*, *155*(6), 1282–95. doi:10.1016/j.cell.2013.10.054

Callender, C., & Cohen, J. (2006). There Is No Special Problem About Scientific Representation. *THEORIA. An International Journal for Theory, History and Foundations of Science*, *21*(1), 67–85. doi:10.1387/THEORIA.554

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press. doi:10.1093/0198247044.001.0001

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Cassa Macedo, A., Oliveira Vilela de Faria, A., & Ghezzi, P. (2019). Boosting the Immune System, From Science to Myth: Analysis the Infosphere With Google. *Frontiers in Medicine*, *6*, 165. doi:10.3389/fmed.2019.00165

Cauchemez, S., Bhattarai, A., Marchbanks, T. L., Fagan, R. P., Ostroff, S., Ferguson, N. M., et al. (2011). Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proceedings of the National Academy of Sciences*, *108*(7), 2825–2830. doi:10.1073/PNAS.1008895108

CDC. (2021). People with Certain Medical Conditions. https://www.cdc.gov/coronavirus/2019-

ncov/need-extra-precautions/people-with-medical-conditions.html. Accessed 29 March 2021

Chakravartty, A. (2010). Informational versus functional theories of scientific representation. *Synthese*, *172*(2), 197–213. doi:10.1007/s11229-009-9502-3

Chang, S. L., Harding, N., Zachreson, C., Cliff, O. M., & Prokopenko, M. (2020). Modelling transmission and control of the COVID-19 pandemic in Australia. *Nature Communications*, *11*(1), 1–13. doi:10.1038/s41467-020-19393-6

Chirimuuta, M. (2014). Minimal models and canonical neural computations: the distinctness of computational explanation in neuroscience. *Synthese*, *191*(2), 127–153. doi:10.1007/s11229-013-0369-y

Chiu, L., Bazin, T., Truchetet, M. E., Schaeverbeke, T., Delhaes, L., & Pradeu, T. (2017). Protective microbiota: From localized to long-reaching co-immunity. *Frontiers in Immunology*, *8*(DEC), 1. doi:10.3389/fimmu.2017.01678

Clark, W. R. (2007). *In Defense of Self: How the Immune System Really Works in Managing Health and Disease*. Oxford: Oxford University Press.

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine*, *57*(6), 745–747. doi:10.1016/J.YPMED.2012.10.020

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the Evidence Hierarchy. *Topoi*, *33*(2), 339–360. doi:10.1007/s11245-013-9220-9

Cliff, O. M., Harding, N., Piraveenan, M., Erten, E. Y., Gambhir, M., & Prokopenko, M. (2018). Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: An agent-based modelling approach. *Simulation Modelling Practice and Theory*, *87*, 412–431. doi:10.1016/J.SIMPAT.2018.07.005

Condamine, T., Ramachandran, I., Youn, J.-I., & Gabrilovich, D. I. (2015). Regulation of Tumor Metastasis by Myeloid-Derived Suppressor Cells. *Annual Review of Medicine*, *66*(1), 97–110. doi:10.1146/annurev-med-051013-052304

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative Reasoning. *Philosophy of Science*, *74*(1), 48–68. doi:10.1086/519478

Coronavirus update for Victoria. (2020, June 22). https://www.dhhs.vic.gov.au/coronavirus-update-victoria-22-june-2020

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376. doi:10.1007/s11229-006-9097-x

Craver, C. F., & Darden, L. (2013). *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago: Chicago University Press.

Davies, D. M. (2018). *The Beautiful Cure: The Revolution in Immunology and What It Means for Your Health*. Chicago: Chicago University Press.

Davis, M. M. (2020). Systems immunology. *Current Opinion in Immunology*, *65*, 79–82. doi:10.1016/j.coi.2020.06.006

Decker, W. K., da Silva, R. F., Sanabria, M. H., Angelo, L. S., Guimarães, F., Burt, B. M., et al. (2017). Cancer Immunotherapy: Historical Perspective of a Clinical Revolution and Emerging Preclinical Animal Models. *Frontiers in Immunology*, *8*, 829. doi:10.3389/fimmu.2017.00829

Delgado, J. C., Baena, A., Thim, S., & Goldfeld, A. E. (2002). Ethnic-Specific Genetic Associations with Pulmonary Tuberculosis. *The Journal of Infectious Diseases*, *186*(10), 1463–1468.

doi:10.1086/344891

Deulofeu, R., Suárez, J., & Pérez-Cervera, A. (2019). Explaining the behaviour of random ecological networks: the stability of the microbiome as a case of integrative pluralism. *Synthese*, 1–23. doi:10.1007/s11229-019-02187-9

Dhabhar, F. S. (2014). Effects of stress on immune function: The good, the bad, and the beautiful. *Immunologic Research*, *58*(2–3), 193–210. doi:10.1007/s12026-014-8517-0

Dietrich, M. R., Ankeny, R. A., Crowe, N., Green, S., & Leonelli, S. (2020). How to choose your research organism. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *80*, 101227. doi:10.1016/J.SHPSC.2019.101227

Dignum, F., Dignum, V., Davidsson, P., Ghorbani, A., van der Hurk, M., Jensen, M., et al. (2020). Analysing the combined health, social and economic impacts of the corovanvirus pandemic using agent-based social simulation. http://arxiv.org/abs/2004.12809. Accessed 16 May 2020

Dobosz, P., & Dzieciątkowski, T. (2019). The Intriguing History of Cancer Immunotherapy. *Frontiers in Immunology*, *10*, 2965. doi:10.3389/fimmu.2019.02965

Druzd, D., Matveeva, O., Ince, L., Harrison, U., He, W., Schmal, C., et al. (2017). Lymphocyte Circadian Clocks Control Lymph Node Trafficking and Adaptive Immune Responses. *Immunity*, *46*(1), 120–132. doi:10.1016/j.immuni.2016.12.011

Dunbar, K. N. (2002). Understanding the role of cognition in science: The Science as Category framework. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 154–170). Cambridge: Cambridge University Press.

Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., & Schreiber, R. D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. *Nature Immunology*, *3*(11), 991–998. doi:10.1038/ni1102-991

Eberl, G. (2010). A new vision of immunity: Homeostasis of the superorganism. *Mucosal Immunology*, *3*(5), 450–460. doi:10.1038/mi.2010.20

Eberl, G. (2016). Immunity by equilibrium. *Nature Reviews Immunology*, *16*(8), 524–532. doi:10.1038/nri.2016.75

Eberl, G., & Pradeu, T. (2018). Towards a General Theory of Immunity? *Trends in Immunology*, *39*(4), 261–263. doi:10.1016/j.it.2017.11.004

Eisinger, D., & Thulke, H.-H. (2008). Spatial pattern formation facilitates eradication of infectious diseases. *Journal of Applied Ecology*, *45*(2), 415–423. doi:10.1111/j.1365-2664.2007.01439.x

Elgin, C. (2007). Understanding and the facts. *Philosophical Studies*, *132*(1), 33–42. doi:10.1007/s11098-006-9054-z

Elgin, C. (2017). *True enough*. Cambridge (Mass.): MIT Press.

Enríquez, J. A. (2019). Mind your mouse strain. *Nature Metabolism*, *1*(1), 5–7. doi:10.1038/s42255-018-0018-3

Eubank, S., Eckstrand, I., Lewis, B., Venkatramanan, S., Marathe, M., & Barrett, C. L. (2020). Commentary on Ferguson, et al., "Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand." *Bulletin of Mathematical Biology*, *82*(4). doi:10.1007/s11538-020-00726-x

Everhart, J. E., Kruszon-Moran, D., Perez-Perez, G. I., Tralka, T. S., & McQuillan, G. (2000). Seroprevalence and Ethnic Differences in Helicobacter pylori Infection among Adults in the

United States. *The Journal of Infectious Diseases*, *181*(4), 1359–1363. doi:10.1086/315384

Fagan, M. B. (2016). Generative models: Human embryonic stem cells and multiple modeling relations. *Studies in History and Philosophy of Science Part A*, *56*, 122–134. doi:10.1016/j.shpsa.2015.10.003

Fajgenbaum, D. C., & June, C. H. (2020). Cytokine Storm. *New England Journal of Medicine*, *383*(23), 2255–2273. doi:10.1056/NEJMra2026131

Farkona, S., Diamandis, E. P., & Blasutig, I. M. (2016). Cancer immunotherapy: the beginning of the end of cancer? *BMC Medicine*, *14*(1). doi:10.1186/s12916-016-0623-5

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., et al. (2020). Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. London: Imperial College London. https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf

Fidler, I. J. (2003). The pathogenesis of cancer metastasis: the “Seed and soil” hypothesis revisited. *Nature Reviews Cancer*, *3*(6), 453–458. doi:10.1038/nrc1098

Files, J. K., Boppana, S., Perez, M. D., Sarkar, S., Lowman, K. E., Qin, K., et al. (2021). Sustained cellular immune dysregulation in individuals recovering from SARS-CoV-2 infection. *Journal of Clinical Investigation*, *131*(1). doi:10.1172/JCI140491

French, S. (2003). A Model-Theoretic Account of Representation (Or, I Don’t Know Much about Art…but I Know It Involves Isomorphism). *Philosophy of Science*, *70*(5), 1472–1483. doi:10.1086/377423

Frese, K. K., & Tuveson, D. A. (2007). Maximizing mouse cancer models. *Nature Reviews Cancer*, *7*(9), 654–658. doi:10.1038/nrc2192

Frieden, T. R., & Lee, C. T. (2020). Identifying and Interrupting Superspreading Events—Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, *26*(6). doi:10.3201/eid2606.200495

Frigg, R. (2006). Scientific Representation and the Semantic View of Theories. *THEORIA. An International Journal for Theory, History and Foundations of Science*, *21*(1), 49–65. doi:10.1387/THEORIA.553

Frigg, R. (2010). Models and fiction. *Synthese*, *172*(2), 251–268. doi:10.1007/s11229-009-9505-0

Frigg, R., & Hartmann, S. (2020). Models in science. In (E. N. Zalta, Ed.)*The Stanford Encyclopedia of Philosophy (Spring 2020 Edition)*. https://plato.stanford.edu/archives/sum2018/entries/models-science/

Frigg, R., & Nguyen, J. (2016). The Fiction View of Models Reloaded. *The Monist*, *99*(3), 225–242. doi:10.1093/monist/onw002

Frigg, R., & Nguyen, J. (2017). Models and Representation. In L. Magnani & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 49–102). Dordrecht: Springer. doi:10.1007/978-3-319-30526-4_3

Frigg, R., & Nguyen, J. (2018). The turn of the valve: representing with material models. *European Journal for Philosophy of Science*, *8*(2), 205–224. doi:10.1007/s13194-017-0182-4

Frigg, R., & Nguyen, J. (2020). *Modelling Nature: An Opinionated Introduction to the Scientific Representation*. Springer.

Fu, X., Ying, Q., Zeng, T., Long, T., & Wang, Y. (2020). Simulating and forecasting the cumulative confirmed cases of SARS-CoV-2 in China by Boltzmann function-based regression analyses. *Journal of Infection*, *80*(5), 578–606. doi:10.1016/J.JINF.2020.02.019

Fumagalli, R. (2015). No Learning from Minimal Models. *Philosophy of Science*, *82*(5), 798–809. doi:10.1086/683281

Gabrilovich, D. I., & Nagaraj, S. (2009). Myeloid-derived suppressor cells as regulators of the immune system. *Nature Reviews Immunology*, *9*(3), 162–174. doi:10.1038/nri2506

Garner, H., & de Visser, K. E. (2020). Immune crosstalk in cancer progression and metastatic spread: a complex conversation. *Nature Reviews Immunology*, *20*(8), 483–497. doi:10.1038/s41577-019-0271-z

Gelfert, A. (2016). *How to Do Science with Models*. Dordrecht: Springer. doi:10.1007/978-3-319-27954-1

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

Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, *71*(5), 742–752. doi:10.1086/425063

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, *172*(2), 269–281. doi:10.1007/s11229-009-9506-z

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., et al. (2020). A SIDARTHE Model of COVID-19 Epidemic in Italy. http://arxiv.org/abs/2003.09861. Accessed 16 May 2020

Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 443–464. doi:10.1016/J.SHPSC.2005.03.011

Glennan, S. (2017). *The new mechanical philosophy*. Oxford: Oxford University Press.

Glennan, S., & Illari, P. M. (2018). Varieties of mechanisms. In S. Glennan & P. M. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 91–103). London: Routledge.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology & Philosophy*, *21*(5), 725–740. doi:10.1007/s10539-006-9054-6

Godfrey-Smith, P. (2009). Abstractions, Idealizations, and Evolutionary Biology. In A. Barberousse, M. Morange, & T. Pradeu (Eds.), *Mapping the Future of Biology* (pp. 47–56). Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-9636-5_4

Goodman, N. (1976). *Languages of Art* (2nd ed.). Indianapolis: Hackett Publishing Company.

Green, S., Dam, M. S., & Svendsen, M. N. (2021). Mouse avatars of human cancers: the temporality of translation in precision oncology. *History and Philosophy of the Life Sciences*, *43*(1), 27. doi:10.1007/s40656-021-00383-w

Green, S., Dietrich, M. R., Leonelli, S., & Ankeny, R. A. (2018). 'Extreme' organisms and the problem of generalization: interpreting the Krogh principle. *History and Philosophy of the Life Sciences*, *40*. doi:10.1007/s40656-018-0231-0

Greslehner, G. P. (2020). Not by structures alone: Can the immune system recognize microbial functions? *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *84*, 101336. doi:10.1016/j.shpsc.2020.101336

Guarda, G., Braun, M., Staehli, F., Tardivel, A., Mattmann, C., Förster, I., et al. (2011). Type I Interferon Inhibits Interleukin-1 Production and Inflammasome Activation. *Immunity*, *34*(2), 213–223. doi:10.1016/j.immuni.2011.02.006

Halina, M. (2018). Mechanistic explanation and its limits. In S. Glennan & P. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 213–224). New York: Routledge. doi:10.4324/9781315731544-16

Hardesty, R. A. (2018). Much ado about mice: Standard-setting in model organism research. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *68–69*, 15–24. doi:10.1016/J.SHPSC.2018.04.001

Hernan, M. A. (2015). Invited Commentary: Agent-Based Models for Causal Inference--Reweighting Data and Theory in Epidemiology. *American Journal of Epidemiology*, *181*(2), 103–105. doi:10.1093/aje/kwu272

Holmes, F. L. (1993). The old martyr of science: The frog in experimental physiology. *Journal of the History of Biology*, *26*(2), 311–328. doi:10.1007/BF01061972

Huber, L., & Keuck, L. K. (2013). Mutant mice: Experimental organisms as materialised models in biomedicine. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *44*(3), 385–391. doi:10.1016/J.SHPSC.2013.03.001

Hughes, R. I. G. (1997). Models and Representation. *Philosophy of Science*, *64*, S325–S336. doi:10.1086/392611

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119–135. doi:10.1007/s13194-011-0038-2

Institute of Medicine. (2006). *Ending the War Metaphor: The Changing Agenda for Unraveling the Host-Microbe Relationship*. Washington, DC: National Academies Press. doi:10.17226/11669

Irving, A. T., Ahn, M., Goh, G., Anderson, D. E., & Wang, L. F. (2021). Lessons from the host defences of bats, a unique viral reservoir. *Nature*, *589*(7842), 363–370. doi:10.1038/s41586-020-03128-0

Jain, A., Chaurasia, R., Sengar, N. S., Singh, M., Mahor, S., & Narain, S. (2020). Analysis of vitamin D level among asymptomatic and critically ill COVID-19 patients and its correlation with inflammatory markers. *Scientific Reports*, *10*(1), 1–8. doi:10.1038/s41598-020-77093-z

Jebeile, J. (2017). Idealizations in empirical modeling. In J. Lenhard & M. Carrier (Eds.), *Mathematics as a Tool:Tracing new roles of mathematics in the sciences* (pp. 213–232). Dordrecht: Springer.

Jebeile, J., & Kennedy, A. G. (2015). Explaining with Models: The Role of Idealizations. *International Studies in the Philosophy of Science*, *29*(4), 383–392. doi:10.1080/02698595.2015.1195143

Jones, M. R. (2005). Idealization and abstraction: A framework. In M. R. Jones & N. Cartwright (Eds.), *Idealization XII: Correcting the model* (pp. 173–217). Amsterdam: Rodopi.

Jones, N. (2018). Strategies of Explanatory Abstraction in Molecular Systems Biology. *Philosophy of Science*, *85*(5), 955–968. doi:10.1086/699742

Julia, V., McSorley, S. S., Malherbe, L., Breittmayer, J.-P., Girard-Pipau, F., Beck, A., & Glaichenhaus, N. (2000). Priming by Microbial Antigens from the Intestinal Flora Determines the Ability of CD4 + T Cells to Rapidly Secrete IL-4 in BALB/c Mice Infected with Leishmania major. *The Journal of Immunology*, *165*(10), 5637–5645. doi:10.4049/jimmunol.165.10.5637

Kampourakis, K. (2020). Why Does It Matter That Many Biology Concepts Are Metaphors? In K. Kampourakis & T. Uller (Eds.), *Philosophy of Science for Biologists* (pp. 102–122). Cambridge: Cambridge University Press.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339–373. doi:10.1007/s11229-011-9970-0

Kaplan, D. M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science*, *78*(4), 601–627. doi:10.1086/661755

Karp, P., & Davey, M. (2020, March 16). Why Australia is not shutting schools to help control the spread of coronavirus. *The Guardian*. https://www.theguardian.com/world/2020/mar/16/why-australia-is-not-shutting-schools-to-help-control-the-spread-of-coronavirus

Katoh, H., Wang, D., Daikoku, T., Sun, H., Dey, S. K., & Dubois, R. N. (2013). CXCR2-expressing myeloid-derived suppressor cells are essential to promote colitis-associated tumorigenesis. *Cancer cell*, *24*(5), 631–44. doi:10.1016/j.ccr.2013.10.009

Keller, M., Mazuch, J., Abraham, U., Eom, G. D., Herzog, E. D., Volk, H. D., et al. (2009). A circadian clock in macrophages controls inflammatory immune responses. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21407–21412. doi:10.1073/pnas.0906361106

Kelly, M. P., Kelly, R. S., & Russo, F. (2014). The Integration of Social, Behavioral, and Biological Mechanisms in Models of Pathogenesis. *Perspectives in Biology and Medicine*, *57*(3), 308–328. doi:10.1353/pbm.2014.0026

Kemp, S. A., Collier, D. A., Datir, R. P., Ferreira, I. A. T. M., Gayed, S., Jahun, A., et al. (2021). SARS-CoV-2 evolution during treatment of chronic infection. *Nature*, *592*(7853), 277–282. doi:10.1038/s41586-021-03291-y

Kennedy, A. G. (2012). A non representationalist view of model explanation. *Studies in History and Philosophy of Science Part A*, *43*(2), 326–332. doi:10.1016/J.SHPSA.2011.12.029

Khosrowi, D. (2020). Getting Serious about Shared Features. *The British Journal for the Philosophy of Science*, *71*(2), 523–546. doi:10.1093/bjps/axy029

Kissler, S. M., Tedijanto, C., Lipsitch, M., & Grad, Y. (2020). Social distancing strategies for curbing the COVID-19 epidemic. *medRxiv*, 2020.03.22.20041079. doi:10.1101/2020.03.22.20041079

Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics*, *5*(11), 826–837. doi:10.1038/nrg1471

Knaus, C., Wahlquist, C., & Remeikis, A. (2020, March 22). PM announces pubs, clubs and cinemas to close, schools stay open in stage one measures – as it happened. *The Guardian*. https://www.theguardian.com/world/live/2020/mar/22/coronavirus-updates-live-australia-nsw-victoria-qld-tasmania-cases-government-stimulus-latest-update-news

Knuuttila, T. (2005). Models, Representation, and Mediation. *Philosophy of Science*, *72*(5), 1260–1271. doi:10.1086/508124

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, *42*(2), 262–271. doi:10.1016/j.shpsa.2010.11.034

Knuuttila, T. (2017). Imagination extended and embedded: artifactual versus fictional accounts of models. *Synthese*, 1–21. doi:10.1007/s11229-017-1545-2

Knuuttila, T., & Loettgers, A. (2017). Modelling as Indirect Representation? The Lotka–Volterra Model Revisited. *The British Journal for the Philosophy of Science*, *68*(4), 1007–1036. doi:10.1093/bjps/axv055

Koropatnick, T. A., Engle, J. T., Apicella, M. A., Stabb, E. V., Goldman, W. E., & McFall-Ngai, M. J. (2004). Microbial factor-mediated development in a host-bacterial mutualism. *Science*, *306*(5699), 1186–1188. doi:10.1126/science.1102218

Krizhanovsky, V., Yon, M., Dickins, R. A., Hearn, S., Simon, J., Miething, C., et al. (2008). Senescence of Activated Stellate Cells Limits Liver Fibrosis. *Cell*, *134*(4), 657–667. doi:10.1016/j.cell.2008.06.049

Kumar, V., Patel, S., Tcyganov, E., & Gabrilovich, D. I. (2016). The Nature of Myeloid-Derived Suppressor Cells in the Tumor Microenvironment. *Trends in immunology*, *37*(3), 208–220. doi:10.1016/j.it.2016.01.004

La Caze, A. (2009). Evidence-Based Medicine Must Be ... *Journal of Medicine and Philosophy*, *34*(5), 509–527. doi:10.1093/jmp/jhp034

La Caze, A. (2011). The role of basic science in evidence-based medicine. *Biology & Philosophy*, *26*(1), 81–98. doi:10.1007/s10539-010-9231-5

Lai, A., Bergna, A., Acciarri, C., Galli, M., & Zehender, G. (2020). Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *Journal of Medical Virology*, *92*(6), 675–679. doi:10.1002/jmv.25723

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: Chicago University Press.

Laplane, L. (2016). *Cancer Stem Cells: Philosophy and Therapies*. Cambridge (Mass.): Harvard University Press.

Laplane, L., Duluc, D., Bikfalvi, A., Larmonier, N., & Pradeu, T. (2019). Beyond the tumour microenvironment. *International Journal of Cancer*, *145*(10), 2611–2618. doi:10.1002/ijc.32343

Laplane, L., Duluc, D., Larmonier, N., Pradeu, T., & Bikfalvi, A. (2018). The Multiple Layers of the Tumor Environment. *Trends in cancer*, *4*(12), 802–809. doi:10.1016/j.trecan.2018.10.002

Laplane, L., Mantovani, P., Adolphs, R., Chang, H., Mantovani, A., McFall-Ngai, M., et al. (2019). Why science needs philosophy. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 3948–3952. doi:10.1073/pnas.1900357116

Laurent, P., Jolivel, V., Manicki, P., Chiu, L., Contin-Bordes, C., Truchetet, M. E., & Pradeu, T. (2017). Immune-mediated repair: A matter of plasticity. *Frontiers in Immunology*, *8*(APR), 24. doi:10.3389/fimmu.2017.00454

Lazarus, R., Vercelli, D., Palmer, L. J., Klimecki, W. J., Silverman, E. K., Richter, B., et al. (2002). Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunological Reviews*, *190*(1), 9–25. doi:10.1034/j.1600-065X.2002.19002.x

Leach, D. R., Krummel, M. F., & Allison, J. P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. *Science (New York, N.Y.)*, *271*(5256), 1734–6. doi:10.1126/science.271.5256.1734

Lee, S., Kim, T., Lee, E., Lee, C., Kim, H., Rhee, H., et al. (2020). Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. *JAMA Internal Medicine*, *180*(11), 1447–1452. doi:10.1001/jamainternmed.2020.3862

Lemoine, M. (2017). Animal extrapolation in preclinical studies: An analysis of the tragic case of TGN1412. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *61*, 35–45. doi:10.1016/J.SHPSC.2016.12.004

Leonelli, S. (2019). What distinguishes data from models? *European Journal for Philosophy of Science*, *9*. doi:10.1007/s13194-018-0246-0

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147–151. doi:10.1038/nature01763

Levy, A. (2012). Models, Fictions, and Realism: Two Packages. *Philosophy of Science*, *79*(5), 738–748. doi:10.1086/667992

Levy, A. (2013). Three kinds of new mechanism. *Biology & Philosophy*, *28*(1), 99–114. doi:10.1007/s10539-012-9337-z

Levy, A. (2015). Modeling without models. *Philosophical Studies*, *172*(3), 781–798. doi:10.1007/s11098-014-0333-9

Levy, A. (2018). Idealization and abstraction: refining the distinction. *Synthese*, 1–18. doi:10.1007/s11229-018-1721-z

Levy, A., & Bechtel, W. (2013). Abstraction and the Organization of Mechanisms. *Philosophy of Science*, *80*(2), 241–261. doi:10.1086/670300

Levy, A., & Currie, A. (2015). Model Organisms are Not (Theoretical) Models. *The British Journal for the Philosophy of Science*, *66*(2), 327–348. doi:10.1093/bjps/axt055

Lewis, D. (2020). Is the coronavirus airborne? Experts can't agree. *Nature*, *580*(7802), 175–175. doi:10.1038/d41586-020-00974-w

Li, C., Ji, F., Wang, L., Wang, L., Hao, J., Dai, M., et al. (2020). Asymptomatic and Human-to-Human Transmission of SARS-CoV-2 in a 2-Family Cluster, Xuzhou, China. *Emerging Infectious Diseases*, *26*(7). doi:10.3201/eid2607.200718

Liang, G., & Bushman, F. D. (2021). The human virome: assembly, composition and host interactions. *Nature Reviews Microbiology*, 1–14. doi:10.1038/s41579-021-00536-5

Liu, Y., & Cao, X. (2016). Characteristics and Significance of the Pre-metastatic Niche. *Cancer Cell*, *30*(5), 668–681. doi:10.1016/J.CCELL.2016.09.011

Liu, Y., Eggo, R. M., & Kucharski, A. J. (2020). Secondary attack rate and superspreading events for SARS-CoV-2. *The Lancet*, *395*(10227), e47. doi:10.1016/S0140-6736(20)30462-1

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

Lochmiller, R. L., & Deerenberg, C. (2000). Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos*, *88*(1), 87–98. doi:10.1034/j.1600-0706.2000.880110.x

Love, A. C., & Nathan, M. J. (2015). The Idealization of Causation in Mechanistic Explanation. *Philosophy of Science*, *82*(5), 761–774. doi:10.1086/683263

Löwy, I. (1996). Metaphors of immunology: war and peace. *História, ciências, saúde--Manguinhos*, *3*(1), 7–23. doi:10.1590/S0104-59701996000100002

Lucas, C., Wong, P., Klein, J., Castro, T. B. R., Silva, J., Sundaram, M., et al. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*, *584*(7821), 463–469.

doi:10.1038/s41586-020-2588-y

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25. doi:10.1086/392759

Magnani, L., & Bertolotti, T. (Eds.). (2017). *Springer Handbook of Model-Based Science*. Dordrecht: Springer. doi:10.1007/978-3-319-30526-4

Mäki, U. (1992). On the method of isolation in economics. In C. Dilworth (Ed.), *Idealization IV: Intelligibility in Science* (pp. 319–354). Amsterdam: Rodopi.

Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, *12*(2), 303–315. doi:10.1080/13501780500086255

Mäki, U. (2009). MISSing the World. Models as Isolations and Credible Surrogate Systems. *Erkenntnis*, *70*(1), 29–43. doi:10.1007/s10670-008-9135-9

Mäki, U. (2011). Models and the locus of their truth. *Synthese*, *180*(1), 47–63. doi:10.1007/s11229-009-9566-0

Mantovani, A., Cassatella, M. A., Costantini, C., & Jaillon, S. (2011). Neutrophils in the activation and regulation of innate and adaptive immunity. *Nature Reviews Immunology*, *11*(8), 519–531. doi:10.1038/nri3024

Marshall, B. D. L., & 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

Matzinger, P. (1994). Tolerance, Danger, and the Extended Family. *Annual Review of Immunology*, *12*(1), 991–1045. doi:10.1146/annurev.iy.12.040194.005015

Matzinger, P. (2002). The danger model: A renewed sense of self. *Science*, *296*(5566), 301–305. doi:10.1126/science.1071059

Maugeri, P., & Blasimme, A. (2011). Humanised models of cancer in molecular medicine: the experimental control of disanalogy. *History and philosophy of the life sciences*, *33*(4), 603–21. https://pubmed.ncbi.nlm.nih.gov/22662512/

Maziarz, M. (2020). *The Philosophy of Causality in Economics: Causal Inferences and Policy Proposals*. New York: Routledge.

Medzhitov, R., & Janeway, C. A. (2002). Decoding the patterns of self and nonself by the innate immune system. *Science*, *296*(5566), 298–300. doi:10.1126/science.1068883

Medzhitov, R., Schneider, D. S., & Soares, M. P. (2012). Disease tolerance as a defense strategy. *Science*, *335*(6071), 936–941. doi:10.1126/science.1214935

Mellman, I., Coukos, G., & Dranoff, G. (2011). Cancer immunotherapy comes of age. *Nature*, *480*(7378), 480–489. doi:10.1038/nature10673

Méthot, P.-O., & Alizon, S. (2014). What is a pathogen? Toward a process view of host-parasite interactions. *Virulence*, *5*(8), 775–785. doi:10.4161/21505594.2014.960726

Mitchell, S. D., & Gronenborn, A. M. (2017). After Fifty Years, Why Are Protein X-ray Crystallographers Still in Business? *The British Journal for the Philosophy of Science*, *68*(3), 703–723. doi:10.1093/bjps/axv051

Morgan, M. S. (2012). *The World in the Model: How Economists Work and Think*. Cambridge: Cambridge University Press.

Morgan, M. S., & Morrison, M. (Eds.). (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Morrison, M. (1998). Modelling Nature: Between Physics and the Physical World. *Philosophia Naturalis*, *35*(1), 65–85.

Morrison, M. (2011). One phenomenon, many models: Inconsistency and complementarity. *Studies in History and Philosophy of Science Part A*, *42*(2), 342–351. doi:10.1016/J.SHPSA.2010.11.042

Morrison, M. (2015). *Reconstructing Reality: Models, Mathematics, and Simulations*. Oxford: Oxford University Press.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators* (pp. 10–37). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511660108.003

Morton, J. J., Bird, G., Refaeli, Y., & Jimeno, A. (2016). Humanized Mouse Xenograft Models: Narrowing the Tumor-Microenvironment Gap. *Cancer research*, *76*(21), 6153–6158. doi:10.1158/0008-5472.CAN-16-1260

Muriuki, J. M., Mentzer, A. J., Kimita, W., Ndungu, F. M., Macharia, A. W., Webb, E. L., et al. (2019). Iron Status and Associated Malaria Risk Among African Children. *Clinical Infectious Diseases*, *68*(11), 1807–1814. doi:10.1093/cid/ciy791

Murphy, K., & Weaver, C. (2017). *Janeway's Immunobiology* (9th ed.). New York: Garland Science.

Murray, P. J., & Wynn, T. A. (2011). Protective and pathogenic functions of macrophage subsets. *Nature Reviews Immunology*, *11*(11), 723–737. doi:10.1038/nri3073

Nagaraj, S., Schrum, A. G., Cho, H.-I., Celis, E., & Gabrilovich, D. I. (2010). Mechanism of T cell tolerance induced by myeloid-derived suppressor cells. *Journal of immunology*, *184*(6), 3106–16. doi:10.4049/jimmunol.0902661

Nathan, M. J. (2014). Causation by Concentration. *The British Journal for the Philosophy of Science*, *65*(2), 191–212. doi:10.1093/bjps/axr056

Nation, M., Fortney, T., & Wandersman, A. (2010). Race, Place, and Neighboring: Social Ties among Neighbors in Urban, Suburban, and Rural Contexts. *Environment and Behavior*, *42*(5), 581–596. doi:10.1177/0013916508328599

National Cabinet Statement. (2020, March 29). *Media statement*. https://www.pm.gov.au/media/national-cabinet-statement

National Institute for Health and Care Excellence. (2014). Developing NICE guidelines: the manual. [manual]. London: National Institute for Health and Care Excellence. https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf

Nature Reviews Immunology Editorial. (2021). Immunology shines in the dark. *Nature Reviews Immunology*, *21*(4), 193–194. doi:10.1038/s41577-021-00527-w

Neher, R. A., Dyrdak, R., Druelle, V., Hodcroft, E. B., & Albert, J. (2020). Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. *Swiss Medical Weekly*, *150*(1112). doi:10.4414/smw.2020.20224

Netea, M. G., Domínguez-Andrés, J., Barreiro, L. B., Chavakis, T., Divangahi, M., Fuchs, E., et al. (2020). Defining trained immunity and its role in health and disease. *Nature Reviews Immunology*, *20*(6), 375–388. doi:10.1038/s41577-020-0285-6

Nicholson, D. J. (2012). The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 152–163. doi:10.1016/J.SHPSC.2011.05.014

Norton, J. D. (2012). Approximation and Idealization: Why the Difference Matters. *Philosophy of Science*, *79*(2), 207–232. doi:10.1086/664746

OCEBM Levels of Evidence. (2009). The Oxford 2009 Levels of Evidence. [manual] Oxford Centre for Evidence-Based Medicine. https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/

Odenbaugh, J. (2015). Semblance or similarity? Reflections on Simulation and Similarity. *Biology & Philosophy*, *30*(2), 277–291. doi:10.1007/s10539-014-9446-y

Ouzounova, M., Lee, E., Piranlioglu, R., El Andaloussi, A., Kolhe, R., Demirci, M. F., et al. (2017). Monocytic and granulocytic myeloid derived suppressor cells differentially regulate spatiotemporal tumour plasticity during metastatic cascade. *Nature Communications*, *8*(1). doi:10.1038/ncomms14979

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., et al. (2020). Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*. doi:10.1001/jama.2020.6130

Pannett, R. (2020, February 1). Australia Restricts Travelers From Mainland China as Virus Impact Spreads. *The Wall Street Journal*. https://www.wsj.com/articles/australia-s-qantas-suspends-china-flights-as-virus-impact-spreads-11580536238

Park, J. H., Cornick, S., Nigro, G., Sevrin, G., Déjardin, F., Smits, R., et al. (2019). Innate immune recognition of a bacterial MAMP leads to conditional activation of pro- or anti-inflammatory responses. *bioRxiv*, 717256. doi:10.1101/717256

Parker, W. S. (2020). Model Evaluation: An Adequacy-for-Purpose View. *Philosophy of Science*, *87*(3), 457–477. doi:10.1086/708691

Parkkinen, V.-P. (2017). Are Model Organisms Theoretical Models? *Disputatio*, *9*(47), 471–498.

Parkkinen, V.-P., Russo, F., & Wallmann, C. (2017). Scientific Disagreement and Evidential Pluralism: Lessons from the Studies on Hypercholesterolemia. *Humana.Mente Journal of Philosophical Studies*, *10*(32), 75–116.

Parkkinen, V.-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P. M., Kelly, M. P., et al. (2018). *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. Springer.

Parkkinen, V.-P., & Williamson, J. (2020). Extrapolating from model organisms in pharmacology. In A. La Caze & B. Osimani (Eds.), *Uncertainty in pharmacology: epistemology, methods, and decisions*. Dordrecht: Springer.

Peng, L., Yang, W., Zhang, D., Zhuge, C., & Hong, L. (2020). Epidemic analysis of COVID-19 in China by dynamical modeling. http://arxiv.org/abs/2002.06563. Accessed 16 May 2020

Pereyra, F., Jia, X., McLaren, P. J., Telenti, A., de Bakker, P. I. W., Walker, B. D., et al. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*, *330*(6010), 1551–1557. doi:10.1126/science.1195271

Piotrowska, M. (2013). From humanized mice to human disease: guiding extrapolation from model to target. *Biology & Philosophy*, *28*(3), 439–455. doi:10.1007/s10539-012-9323-5

Plutynski, A. (2017). Evolutionary Perspectives on Molecular Medicine: Cancer from an Evolutionary

Perspective. In G. Boniolo & M. J. Nathan (Eds.), *Philosophy of Molecular Medicine: Foundational Issues in Research and Practice* (pp. 122–144). New York: Routledge.

Plutynski, A. (2018). *Explaining Cancer: Finding Order in Disorder*. New York: Oxford University Press.

Plutynski, A. (2019). Cancer. In *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)*. https://plato.stanford.edu/archives/sum2019/entries/cancer/

Plutynski, A., & Bertolaso, M. (2018). What and How Do Cancer Systems Biologists Explain? *Philosophy of Science*, *85*(5), 942–954. doi:10.1086/699716

Poon, M. M. L., & Farber, D. L. (2020). The Whole Body as the System in Systems Immunology. *iScience*, *23*(9), 101509. doi:10.1016/j.isci.2020.101509

Portides, D. (2007). The Relation between Idealisation and Approximation in Scientific Model Construction. *Science & Education*, *16*(7–8), 699–724. doi:10.1007/s11191-006-9001-6

Portides, D. (2011). Seeking representations of phenomena: Phenomenological models. *Studies in History and Philosophy of Science Part A*, *42*(2), 334–341. doi:10.1016/J.SHPSA.2010.11.041

Portides, D. (2013). Idealization in Economics Modeling. In Hanne Andersen, D. Dieks, W. Gonzalez, T. Uebel, & G. Wheeler (Eds.), *New Challenges to Philosophy of Science* (pp. 253–263). Dordrecht: Springer.

Portides, D. (2018). Idealization and abstraction in scientific modeling. *Synthese*, 1–23. doi:10.1007/s11229-018-01919-7

Potochnik, A. (2015). The diverse aims of science. *Studies in History and Philosophy of Science Part A*, *53*, 71–80. doi:10.1016/J.SHPSA.2015.05.008

Potochnik, A. (2017). *Idealization and the aims of science*. Chicago: University of Chicago Press.

Poznic, M. (2016). Representation and Similarity: Suárez on Necessary and Sufficient Conditions of Scientific Representation. *Journal for General Philosophy of Science*, *47*(2), 331–347. doi:10.1007/s10838-015-9307-7

Pradeu, T. (2012). *The Limits of the Self: Immunology and Biological Identity*. Oxford: Oxford University Press.

Pradeu, T. (2016). Mutualistic viruses and the heteronomy of life. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *59*, 80–88. doi:10.1016/j.shpsc.2016.02.007

Pradeu, T. (2019). *Philosophy of Immunology*. Cambridge: Cambridge University Press. doi:10.1017/9781108616706

Pradeu, T., Jaeger, S., & Vivier, E. (2013). The speed of change: towards a discontinuity theory of immunity? *Nature Reviews Immunology*, *13*(10), 764–769. doi:10.1038/nri3521

Pradeu, T., & Vivier, E. (2016). The discontinuity theory of immunity. *Science Immunology*, *1*(1), aag0479. doi:10.1126/sciimmunol.aag0479

Psillos, S. (2011). Living with the abstract: realism and models. *Synthese*, *180*(1), 3–17. doi:10.1007/s11229-009-9563-3

Railsback, S. F., & Grimm, V. (2012). *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton: Princeton University Press.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., & Medzhitov, R. (2004). Recognition

of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*, *118*(2), 229–241. doi:10.1016/j.cell.2004.07.002

Rankin, L. C., & Artis, D. (2018). Beyond Host Defense: Emerging Functions of the Immune System in Regulating Complex Tissue Physiology. *Cell*, *173*(3), 554–567. doi:10.1016/j.cell.2018.03.013

Redelings, M. D., Sorvillo, F., & Mascola, L. (2007). Increase in Clostridium difficile–related Mortality Rates, United States, 1999–2004. *Emerging Infectious Diseases*, *13*(9), 1417–1419. doi:10.3201/eid1309.061116

Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology*, *19*(1), 43–62. doi:10.1080/1350178X.2012.661069

Remy, K. E., Mazer, M., Striker, D. A., Ellebedy, A. H., Walton, A. H., Unsinger, J., et al. (2020). Severe immunosuppression and not a cytokine storm characterizes COVID-19 infections. *JCI Insight*, *5*(17), e140329. doi:10.1172/jci.insight.140329

Reutlinger, A., Hangleiter, D., & Hartmann, S. (2018). Understanding (with) Toy Models. *The British Journal for the Philosophy of Science*, *69*(4), 1069–1099. doi:10.1093/bjps/axx005

Ribas, A., & Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science*, *359*(6382), 1350–1355. doi:10.1126/SCIENCE.AAR4060

Rice, C. C. (2015). Moving Beyond Causes: Optimality Models and Scientific Explanation. *Noûs*, *49*(3), 589–615. doi:10.1111/nous.12042

Rice, C. C. (2016). Factive scientific understanding without accurate representation. *Biology & Philosophy*, *31*(1), 81–102. doi:10.1007/s10539-015-9510-2

Rice, C. C. (2018). Idealized models, holistic distortions, and universality. *Synthese*, *195*(6), 2795–2819. doi:10.1007/s11229-017-1357-4

Rice, C. C. (2019). Models Don't Decompose That Way: A Holistic View of Idealized Models. *The British Journal for the Philosophy of Science*, *70*(1), 179–208. doi:10.1093/bjps/axx045

Rich, R. R., Fleisher, T. A., Shearer, W. T., Schroeder, H., Frew, A. J., & Weyand, C. M. (2019). *Clinical Immunology: Principles and Practice* (5th ed.). Elsevier.

Rocca, E. (2018). The judgements that evidence-based medicine adopts. *Journal of Evaluation in Clinical Practice*, *24*(5), 1184–1190. doi:10.1111/jep.12994

Rodier, F., & Campisi, J. (2011). Four faces of cellular senescence. *Journal of Cell Biology*, *192*(4), 547–556. doi:10.1083/jcb.201009094

Rohwer, Y., & Rice, C. C. (2013). Hypothetical Pattern Idealization and Explanatory Models. *Philosophy of Science*, *80*(3), 334–355. doi:10.1086/671399

Rongvaux, A., Willinger, T., Martinek, J., Strowig, T., Gearty, S. V, Teichmann, L. L., et al. (2014). Development and function of human innate immune cells in a humanized mouse model. *Nature Biotechnology*, *32*(4), 364–372. doi:10.1038/nbt.2858

Russo, F., & Williamson, J. (2007). Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science*, *21*(2), 157–170. doi:10.1080/02698590701498084

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-Based Medicine. How to Practice and Teach EBM* (2nd ed.). London: Churchull Livingstone.

Sakaguchi, S. (2006). Regulatory T cells: Meden Agan. *Immunological Reviews*, *212*(1), 5–7. doi:10.1111/j.0105-2896.2006.00425.x

Salis, F. (2016). The Nature of Model-World Comparisons. *The Monist*, *99*(3), 243–259. doi:10.1093/monist/onw003

Salis, F. (2019). The New Fiction View of Models. *The British Journal for the Philosophy of Science*. doi:10.1093/bjps/axz015

Sanmamed, M. F., Chester, C., Melero, I., & Kohrt, H. (2016). Defining the optimal murine models to investigate immune checkpoint blockers and their combination with other immunotherapies. *Annals of Oncology*, *27*(7), 1190–1198. doi:10.1093/ANNONC/MDW041

Sansonetti, P. J., & Medzhitov, R. (2009). Learning Tolerance while Fighting Ignorance. *Cell*, *138*(3), 416–420. doi:10.1016/j.cell.2009.07.024

Schultze, J. L., & Aschenbrenner, A. C. (2021). COVID-19 and the human innate immune system. *Cell*, *184*(7), 1671–1692. doi:10.1016/j.cell.2021.02.029

Shaw, A. C., Joshi, S., Greenwood, H., Panda, A., & Lord, J. M. (2010). Aging of the innate immune system. *Current opinion in immunology*, *22*(4), 507–513. doi:10.1016/j.coi.2010.05.003

Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, *587*(7835), 657–662. doi:10.1038/s41586-020-2601-5

Shultz, L. D., Brehm, M. A., Garcia-Martinez, J. V., & Greiner, D. L. (2012). Humanized mice for immune system investigation: progress, promise and challenges. *Nature Reviews Immunology*, *12*(11), 786–798. doi:10.1038/nri3311

Shultz, L. D., Ishikawa, F., & Greiner, D. L. (2007). Humanized mice in translational biomedical research. *Nature Reviews Immunology*, *7*(2), 118–130. doi:10.1038/nri2017

Skillings, D. J. (2015). Mechanistic Explanation of Biological Processes. *Philosophy of Science*, *82*(5), 1139–1151. doi:10.1086/683446

Smith, K. G. C., & Clatworthy, M. R. (2010). FcγRIIB in autoimmunity and infection: evolutionary and therapeutic implications. *Nature Reviews Immunology*, *10*(5), 328–343. doi:10.1038/nri2762

Squazzoni, F., Polhill, J. G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G., et al. (2020). Computational Models That Matter During a Global Pandemic Outbreak: A Call to Action. *Journal of Artificial Societies and Social Simulation*, *23*(2), 10. doi:10.18564/jasss.4298

Sridhar, D., & Majumder, M. S. (2020). Modelling the pandemic. *BMJ*, *369*. doi:10.1136/BMJ.M1567

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

Stefan, K. L., Kim, M. V., Iwasaki, A., & Kasper, D. L. (2020). Commensal Microbiota Modulation of Natural Resistance to Virus Infection. *Cell*, *183*(5), 1312-1324.e10. doi:10.1016/j.cell.2020.10.047

Steimle, A., Autenrieth, I. B., & Frick, J. S. (2016). Structure and function: Lipid A modifications in commensals and pathogens. *International Journal of Medical Microbiology*, *306*(5), 290–301. doi:10.1016/j.ijmm.2016.03.001

Sterrett, S. G. (2017). Physically Similar Systems - A History of the Concept. In L. Magnani & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 377–411). Dordrecht: Springer. doi:10.1007/978-3-319-30526-4_18

Strevens, M. (2008). *Depth: an account of scientific explanation*. Harvard University Press.

Strevens, M. (2017). How Idealizations Provide Understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science* (pp. 37–49). Routledge. doi:10.4324/9781315686110-10

Stutman, O. (1974). Tumor development after 3-methylcholanthrene in immunologically deficient athymic-nude mice. *Science*, *183*(4124), 534–536. doi:10.1126/science.183.4124.534

Suárez, M. (2003). Scientific representation: against similarity and isomorphism. *International Studies in the Philosophy of Science*, *17*(3), 225–244. doi:10.1080/0269859032000169442

Suárez, M. (2004). An Inferential Conception of Scientific Representation. *Philosophy of Science*, *71*(5), 767–779. doi:10.1086/421415

Suárez, M. (2009). Fictions in scientific practice. In M. Suárez (Ed.), *Fictions in Science: Philosophical Essay on Modeling and Idealization* (pp. 3–15). New York: Routledge.

Suárez, M. (2010). Scientific Representation. *Philosophy Compass*, *5*(1), 91–101. doi:10.1111/j.1747-9991.2009.00261.x

Swiatczak, B., & Tauber, A. I. (2020). Philosophy of Immunology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*. https://plato.stanford.edu/entries/immunology/

Tauber, A. I. (1994). *The Immune Self: Theory or Metaphor?* Cambridge: Cambridge University Press.

Tauber, A. I. (2017). *Immunity: The evolution of an idea*. New York: Oxford University Press.

Tauber, A. I., & Podolsky, S. H. (1994). Frank Macfarlane Burnet and the immune self. *Journal of the History of Biology*, *27*(3), 531–573. doi:10.1007/BF01058996

Teller, P. (2001). Twilight of the Perfect Model Model. *Erkenntnis*, *55*(3), 393–415. doi:10.2307/20013097

Teller, P. (2012). Modeling, Truth, and Philosophy. *Metaphilosophy*, *43*(3), 257–274. doi:10.1111/j.1467-9973.2012.01745.x

The RECOVERY Collaborative Group. (2021). Dexamethasone in Hospitalized Patients with Covid-19. *New England Journal of Medicine*, *384*(8), 693–704. doi:10.1056/NEJMoa2021436

Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors We Think With: The Role of Metaphor in Reasoning. *PLoS ONE*, *6*(2), e16782. doi:10.1371/journal.pone.0016782

Thomasson, A. L. (2020). If Models Were Fictions, Then What Would They Be? In A. Levy & P. Godfrey-Smith (Eds.), *The Scientific Imagination* (pp. 51–74). Oxford: Oxford University Press.

Thomson-Jones, M. (2010). Missing systems and the face value practice. *Synthese*, *172*(2), 283–299. doi:10.1007/s11229-009-9507-y

Thomson-Jones, M. (2012). Modeling without Mathematics. *Philosophy of Science*, *79*(5), 761–772. doi:10.1086/667876

Thomson-Jones, M. (2020). Realism About Missing Systems. In A. Levy & P. Godfrey-Smith (Eds.), *The Scientific Imagination* (pp. 75–101). Oxford: Oxford University Press.

Tobías, A. (2020). Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. *Science of The Total Environment*, *725*, 138539. doi:10.1016/J.SCITOTENV.2020.138539

Toh, B., Wang, X., Keeble, J., Sim, W. J., Khoo, K., Wong, W.-C., et al. (2011). Mesenchymal Transition

and Dissemination of Cancer Cells Is Driven by Myeloid-Derived Suppressor Cells Infiltrating the Primary Tumor. *PLoS Biology*, *9*(9), e1001162. doi:10.1371/journal.pbio.1001162

Toon, A. (2010). The ontology of theoretical modelling: models as make-believe. *Synthese*, *172*(2), 301–315. doi:10.1007/s11229-009-9508-x

Toon, A. (2012a). *Models as Make-Believe*. London: Palgrave Macmillan. doi:10.1057/9781137292230

Toon, A. (2012b). Similarity and Scientific Representation. *International Studies in the Philosophy of Science*, *26*(3), 241–257. doi:10.1080/02698595.2012.731730

Truchetet, M. E., & Pradeu, T. (2018). Re-thinking our understanding of immunity: Robustness in the tissue reconstruction system. *Seminars in Immunology*, *36*, 45–55. doi:10.1016/j.smim.2018.02.013

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, *449*(7164), 804–810. doi:10.1038/nature06244

van Eck, D., & Mennes, J. (2016). Design Explanation and Idealization. *Erkenntnis*, *81*(5), 1051–1071. doi:10.1007/s10670-015-9782-6

Veglia, F., Perego, M., & Gabrilovich, D. (2018). Myeloid-derived suppressor cells coming of age. *Nature Immunology*, *19*(2), 108–119. doi:10.1038/s41590-017-0022-x

Vély, F., Barlogis, V., Vallentin, B., Neven, B., Piperoglou, C., Ebbo, M., et al. (2016). Evidence of innate lymphoid cell redundancy in humans. *Nature Immunology*, *17*(11), 1291–1299. doi:10.1038/ni.3553

Vivier, E., Tomasello, E., Baratin, M., Walzer, T., & Ugolini, S. (2008). Functions of natural killer cells. *Nature Immunology*, *9*(5), 503–510. doi:10.1038/ni1582

Vorms, M. (2011). Representing with imaginary models: Formats matter. *Studies in History and Philosophy of Science Part A*, *42*(2), 287–295. doi:10.1016/j.shpsa.2010.11.036

Walsh, N. C., Kenney, L. L., Jangalwe, S., Aryee, K.-E., Greiner, D. L., Brehm, M. A., & Shultz, L. D. (2017). Humanized Mouse Models of Clinical Disease. *Annual Review of Pathology: Mechanisms of Disease*, *12*(1), 187–215. doi:10.1146/annurev-pathol-052016-100332

Wang, E. Y., Mao, T., Klein, J., Dai, Y., Huck, J. D., Liu, F., et al. (2020). Diverse Functional Autoantibodies in Patients with COVID-19. *medRxiv*, 2020.12.10.20247205. doi:10.1101/2020.12.10.20247205

Wang, Y., Ding, Y., Guo, N., & Wang, S. (2019). MDSCs: Key Criminals of Tumor Pre-metastatic Niche Formation. *Frontiers in Immunology*, *10*, 172. doi:10.3389/fimmu.2019.00172

Wayne, A. (2011). Expanding the Scope of Explanatory Idealization. *Philosophy of Science*, *78*(5), 830–841. doi:10.1086/662277

Weber, M. (2014). Experimental modeling in biology: In vivo representation and stand-ins as modeling strategies. *Philosophy of Science*, *81*(5), 756–769. doi:10.1086/678257

Weinberg, R. A. (2014). *The Biology of Cancer* (2nd ed.). New York: Garland Science.

Weiner, H. L. (2000). Oral tolerance, an active immunologic process mediated by multiple mechanisms. *Journal of Clinical Investigation*, *106*(8), 935–937. doi:10.1172/JCI11348

Weisberg, M. (2004). Qualitative Theory and Chemical Explanation. *Philosophy of Science*, *71*(5), 1071–1081. doi:10.1086/428011

Weisberg, M. (2006). Robustness Analysis. *Philosophy of Science*, *73*(5), 730–742. doi:10.1086/518628

Weisberg, M. (2007). Who is a Modeler? *The British Journal for the Philosophy of Science*, *58*(2), 207–233. doi:10.1093/bjps/axm011

Weisberg, M. (2013). *Simulation and Similarity*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199933662.001.0001

Wilder-Smith, A., & Freedman, D. O. (2020). Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *Journal of Travel Medicine*, *27*(2). doi:10.1093/jtm/taaa020

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Cambridge (Mass.): MIT Press.

Williamson, J. (2019). EBM+: increasing the systematic use of mechanistic evidence. In *Oral Presentations* (Vol. 24, pp. A13–A14). BMJ Publishing Group Ltd. doi:10.1136/bmjebm-2019-EBMLive.25

Wong, G., Liu, W., Liu, Y., Zhou, B., Bi, Y., & Gao, G. F. (2015). MERS, SARS, and Ebola: The Role of Super-Spreaders in Infectious Disease. *Cell Host & Microbe*, *18*(4), 398–401. doi:10.1016/J.CHOM.2015.09.013

Woodward, J. F. (2011). Data and phenomena: a restatement and defense. *Synthese*, *182*(1), 165–179. doi:10.1007/s11229-009-9618-5

Worldometer. (2020). *Australia*. https://www.worldometers.info/coronavirus/country/australia/. Accessed 25 June 2020

Worrall, J. (2010). Evidence: philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice*, *16*(2), 356–362. doi:10.1111/j.1365-2753.2010.01400.x

Worthington, B., & Snape, J. (2020, March 19). Australia blocks arrival of all non-citizens, non-residents in expanded coronavirus travel ban. *ABC News*. https://www.abc.net.au/news/2020-03-19/coronavirus-non-resident-travel-ban-australia/12071640

Wu, J., Dhingra, R., Gambhir, M., & Remais, J. V. (2013). Sensitivity analysis of infectious disease models: methods, advances and their application. *Journal of The Royal Society Interface*, *10*(86), 20121018. doi:10.1098/rsif.2012.1018

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., et al. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174. doi:10.21037/jtd.2020.02.64

Yel, L. (2010). Selective IgA Deficiency. *Journal of Clinical Immunology*, *30*(1), 10–16. doi:10.1007/s10875-009-9357-x

Young, E. (2020). Immunology Is Where Intuition Goes to Die. *The Atlantic*. https://www.theatlantic.com/health/archive/2020/08/covid-19-immunity-is-the-pandemics-central-mystery/614956. Accessed 9 February 2021

Zach, M. (2019). Conceptual Analysis in the Philosophy of Science. *Balkan Journal of Philosophy*, *11*(2), 107–124.

Zitvogel, L., Pitt, J. M., Daillère, R., Smyth, M. J., & Kroemer, G. (2016). Mouse models in oncoimmunology. *Nature Reviews Cancer*, *16*(12), 759–773. doi:10.1038/nrc.2016.91

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., et al. (2020). SARS-CoV-2 Viral Load in

Upper Respiratory Specimens of Infected Patients. *New England Journal of Medicine*, *382*(12), 1177–1179. doi:10.1056/NEJMc2001737

List of abbreviations

ABM = agent-based model

ADR = abstract direct representation

AIDS = acquired immunodeficiency syndrome

DDM = description-driven modeling

EBM = evidence-based medicine

EDM = experimentation-driven modeling

GEMM = genetically engineered mouse model

HIV = human immunodeficiency virus

HLA = human leukocyte antigen

HPSC = hematopoietic stem and progenitor cell

IFN = interferon

ILC = innate lymphoid cell

IPEX = immune dysregulation, polyendocrinopathy, enteropathy, X-linked

MDSC = myeloid-derived suppressor cell

PBMC = peripheral blood mononuclear cell

PDX = patient-derived xenograft

SCID = severe combined immunodeficiency

TLR = toll-like receptor

xGVHD = xenograft versus host disease