MECHANISMS AND MECHANISTIC REASONING IN MEDICINE

Anić, Zvonimir

UNIVERSITY OF RIJEKA
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
DEPARTMENT OF PHILOSOPHY

Zvonimir Anić

MECHANISMS AND MECHANISTIC REASONING IN MEDICINE

DOCTORAL THESIS

Rijeka, 2022.

UNIVERSITY OF RIJEKA
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
DEPARTMENT OF PHILOSOPHY

Zvonimir Anić

MECHANISMS AND MECHANISTIC REASONING IN MEDICINE

DOCTORAL THESIS

Advisor: dr.sc. Davor Pećnjak
Co-advisor: dr.sc. Predrag Šustar

Rijeka, 2022.

SVEUČILIŠTE U RIJECI
FILOZOFSKI FAKULTET
ODSJEK ZA FILOZOFIJU

Zvonimir Anić

MEHANIZMI I MEHANICISTIČKO OBJAŠNJENJE I PREDVIĐANJE U MEDICINI

DOKTORSKI RAD

Mentor: dr.sc. Davor Pećnjak
Komentor: dr.sc. Predrag Šustar

Rijeka, 2022.

Mentor rada: dr.sc. Davor Pećnjak

Komentor rada: dr.sc. Predrag Šustar

Doktorski rad obranjen je dana 12. srpnja 2022. godine u/na Filozofskom fakultetu Sveučilišta u Rijeci, pred povjerenstvom u sastavu:

1. dr.sc. Bojan Borstner
2. dr.sc. Boran Berčić
3. dr.sc. Luca Malatesti
4. _____

ACKNOWLEDGEMENTS

I would like to thank my advisor Davor Pećnjak for his advice, philosophical remarks, support, patience, and friendship.

A special thank you goes to my co-advisor Predrag Šustar. His guidance, support, mentorship, and help made this doctoral thesis far better than it would have been.

For his time, invaluable comments, and advice I have to thank Jonathan Fuller.

Thank you Zdenka, Martina, Aleksandar, and Vito for listening, reading, and commenting. You helped me a lot.

Also, I have to thank all my colleagues at the Institute of philosophy for their support, the Department of Philosophy at the University of Rijeka for their warm welcome when it was needed, and the Department of History and Philosophy of Science at the University of Pittsburgh for their generous offer to host me for 9 months.

Finally, the biggest thank you and gratitude goes to my mom Marijana, my dad Branko, my brother Tomislav, Adrijana, and Tihana. I could not have done this without your support, encouragement, and patience.

SUMMARY

In the late 1990s and early 2000s "The New Mechanistic Philosophy" emerged as a framework for thinking about numerous traditional issues in philosophy of science, but first and foremost, it offered a new account of scientific explanation. The mechanistic account of explanation asserts that the majority of explanations in biological and biomedical sciences are descriptions of biological mechanisms: phenomena are explained by constructing and providing models of mechanisms that are supposed to be causally or constitutively responsible for them. On the other hand, mechanistic reasoning – the mechanistic model of prediction – assumes that knowledge about the inner constitution of mechanisms allows for making predictions about the outcomes of interventions into mechanisms. Within the Evidence-Based Medicine movement in contemporary medicine, however, predictive inferences based on mechanistic evidence (mechanistic reasoning), are considered as low-quality and unreliable evidence. Such a stance on mechanistic evidence and mechanistic reasoning goes against the arguments proposed by mechanistic philosophers over the past 20 years. This dissertation provides a comprehensive discussion on the mechanistic approach to explanation and prediction in order to solve the problem of the unreliability of mechanistic evidence for making prediction claims about the outcomes of medical interventions. Throughout this dissertation I assess three general aspects of the relation between the mechanistic approach to explanation and prediction and contemporary medical science and practice. First, I discuss what makes the mechanistic approach for assessing causal claims in medicine distinct from the epidemiological approach (favored by the Evidence-Based Medicine framework). Second, I discuss the ontological, epistemological, and methodological theses of "The New Mechanistic Philosophy" from the perspective of medical science and practice. Third, I discuss what mechanistic reasoning amounts to and why it so often fails to provide true predictions. Finally, I offer my account of mechanistic reasoning in medicine and criteria for assessing the quality of mechanistic predictions of the outcomes of medical interventions.

Key words: disease causation, mechanistic explanation, mechanistic reasoning, prediction, Evidence-Based Medicine, evidence, medical interventions, dysfunctions.

PROŠIRENI SAŽETAK

Suvremena medicinska znanost i praksa velikim je dijelom obilježena evidencijskim okvirom tzv. medicine temeljene na dokazima (eng. Evidence-Based Medicine). Unutar tog okvira, preferiraju se one tvrdnje o povezanosti rizičnih faktora s određenim bolestima, ishodima medicinskih intervencija te prognostičke i dijagnostičke tvrdnje koje su temeljene na dokazima dobivenim iz epidemioloških studija (randomizirane kontrolirane studije, istraživanja kohorte te istraživanja parova). Dokazi dobiveni iz tih studija odnose se na statističke pojmove poput omjera rizika (eng. risk ratio), omjera izgleda (eng. odds ratio) i pripisivog rizika (eng. attributable risk) te opisuju korelacije između varijabla na populacijskoj razini. S druge strane, unutar filozofije znanosti Nova mehanicistička filozofija u zadnjih se 20 godina nametnula kao dominantan okvir znanstvenog objašnjenja i metodologije u biološkim i biomedicinskim znanostima. Glavna tvrdnja Nove mehanicističke filozofije jest da su znanstvena objašnjenja u biološkim i biomedicinskim znanostima uglavnom strukturirana kao modeli ili opisi bioloških mehanizama koji povezuju određeni uzrok s posljedicom. Nadalje, većina mehanicističkih filozofa prihvaća znanstveni realizam u pogledu tih modela mehanizama: dobri modeli mehanizama opisuju stvarne ontološke strukture. Stoga, prema argumentima mehanicističkih filozofa znanstveno objašnjenje većine bioloških fenomena prihvatljivo je ukoliko je otkriven i detaljno opisan mehanizam koji uzrokuje fenomen. Također, pretpostavlja se da znanje o dijelovima mehanizama i njihovim međusobnim uzročnim i drugim organizacijskim odnosima omogućuje točna predviđanja ishoda intervencija u mehanizme.

Nova mehanicistička filozofija i medicina temeljena na dokazima nude dva vrlo različita poimanja dokaza potrebnih za iskazivanje mnogobrojnih uzročnih tvrdnji u medicini, posebice tvrdnje o ishodima medicinskih intervencija. Cilj ove disertacije je identificirati i razriješiti standardne probleme unutar ontoloških i epistemoloških argumenata Nove mehanicističke filozofije poput odnosa između ontoloških mehanizama i modela mehanizama, kriterija dobrog mehanicističkog objašnjena i strukture mehanicističkog predviđanja te ponuditi odgovore na pitanja odnosa mehanicističke filozofije i medicinske znanosti i prakse poput objašnjenja bolesti unutar mehanicističkog okvira, identifikacije razloga učestalog neuspjeha mehanicističkog predviđanja ishoda medicinskih intervencija te formuliranja kriterija dobrog mehanicističkog predviđanja ishoda medicinskih intervencija. Također, u ovoj disertaciji razmatram i neke do sada neobrađene teme odnosa mehanicističke filozofije i medicinske prakse poput pitanja mehanicističkog pristupa dijagnozi i dijagnostičkim

tvrdnjama. Iz navedenih razloga, ova disertacija predstavlja prvi sveobuhvatni prikaz odnosa između Nove mehanicističke filozofije te suvremene medicinske znanosti i prakse.

Ključne riječi: uzrokovanje bolesti, mehanicističko objašnjenje, predviđanje, medicina temeljena na dokazima, znanstveni dokazi, medicinske intervencije, disfunkcije

CONTENT:

INTRODUCTION

Medicine is both a science and a practice. Whether medical scientists and clinicians seek to explain, predict, or intervene into phenomena, all scientific and practical labor in medicine, in the end, aims to contribute in one way or another to the overall health of individual patients and populations. A commonsensical assumption would have it then that this end cannot be achieved without knowing the causal relations that either maintain normal physiological functions or underlie the onset and progress of diseases. That is, it seems reasonable to presume that if one knows how certain diseases are caused and what kind of processes and mechanisms are involved in their progress, then one should be able to know how to prevent or cure diseases.

Hence, knowing the causes of diseases occupies a central place in both theoretical and clinical aspect of medicine. Theoretical medicine is what students learn when they first enter medical schools. It constitutes the body of knowledge about normal and abnormal bodily conditions, genetic influences, physiological processes, and all of the known biological, social, and psychological factors that influence our physiological and psychological well-being. Theoretical medicine is concerned with ascertaining the causes, correlations and constitutive aspects of our health conditions, illnesses, diseases, and physiological processes. It is concerned with general causal claims and type-explanations (such as that tuberculosis is caused by a specific bacterium called *Mycobacterium tuberculosis* or that obesity presents a major risk condition for different cardiovascular diseases) rather than singular causal claims and token-explanations (such as how did Smith's sepsis come about). Numerous scientists working in different fields of the health sciences, from epidemiology to microbiology, contribute to this research. Concerning causes of diseases, then, the product of this diverse scientific research is systematized into networks of causal pathways that lead to a specific disease. Clinical medicine, on the other hand, is what practitioners are faced with in their daily practice. It is concerned with diagnosis and treatment of individual patients by their doctors. Of course, it is based on theoretical medicine, but it has its own procedures, methodologies, epistemologies, inferences, and ethical considerations.

Whether we are talking about medical science or practice, theoretical or clinical medicine, improving positive patient-relevant outcomes is achieved either by intervention, in terms of treatment or cure, or by prevention. Both prevention and intervention are grounded in knowledge about the causes underlying diseases or normal bodily functions. If smoking causes lung cancer, then refraining from smoking eliminates at least one type of causal process that

leads to lung cancer. Similarly, knowing the actual causal processes that result in cancerous cells enables the formulation of possible medical interventions. Nevertheless, it is far from true that knowledge of the cause or causes of a certain disease will in every case be sufficient to devise treatments or preventions of that disease. That smoking is a cause of lung cancer was already presumed in the 1920s, but the treatments available for patients with lung cancer were limited. Similarly, we can have a fairly good understanding of the cause of some disease but lack any means whatsoever to prevent it (such as for the celiac diseases or Down syndrome).

However, observe that prevention and intervention (as possible ways of achieving positive patient-relevant outcomes) do not aim at the same type of causal processes or causal pathways. That is, refraining from smoking and intervening into the mechanisms responsible for the growth of lung cancer do not aim at the same causal processes or pathways. Therefore, many philosophers and philosophically minded medical scientists distinguish between two types of causal processes (or perhaps, as I will discuss, two types of causal explanations). The first type is that of a typical causal history before the onset of a disease while the other type is involved in the progress of a disease or in causal relations that are constitutive of a certain disease. What does this amount to? In a nutshell, slippery stairs are one of the causes of broken hips, but the mechanism of a fracturing hip bone does not include slippery stairs. Cigarette smoke is one of the causes of high blood pressure, but it does not constitute the mechanism of high blood pressure. This distinction of causal processes is often referred to as a distinction between *etiological* and *pathogenic* processes (the latter sometimes also referred to as the pathophysiological constitution of a disease, e.g., in Dammann 2020). That is, medical scientists and practitioners differentiate between the mechanisms, processes, and causes that lead to disease in the first place and the mechanisms, processes, and causes that obtain after the onset of disease (the latter are constitutive of a disease and they bring about the signs and symptoms of the disease).[1]

Before going further into the discussion, let us first define aforementioned notions of etiology, physiology, pathology, pathophysiology, and pathogenesis.

[1] At this point, perhaps, these should be somewhat clarified. Signs and symptoms, as far as the medical literature is concerned, are differentiated according to who is the observer. A sign is an objective and measurable effect of a disease state measured by the clinician or scientist, while a symptom is a subjective experience of a disease experienced by the diseased.

I take that etiology refers to the exposure factors found to correlate with the onset of a certain disease (e.g., smoking one pack of cigarettes daily for 20 years and lung cancer, or obesity and different cardiovascular diseases) while pathogenesis refers to the biological, chemical, and physical processes that lead to the onset of the disease (e.g., destruction of cilia, accumulation of carcinogenic agents etc.). Although physiology is a term that everyone is familiar with, a clear definition is helpful since two senses of the term are sometimes used interchangeably (especially in the philosophy of medicine literature). In the APS Strategic Plan of the American Physiological Society for 2006-2010 physiology is defined as "the study of the function of organism as integrated systems of molecules, cells, tissues in health and disease" (2006: 163). The popular website webmd.com defines physiology as "the study of how the human body works. It describes the chemistry and physics behind basic body functions, from how molecules behave in cells to how systems of organs work together".[2] As I noted, there is, however, a different sense in which physiology is sometimes used (especially in philosophy in medicine). Physiology in this latter sense refers to the mechanisms themselves, and not the particular field of study or branch of biomedicine and biology. If not stated otherwise, I will use physiology in the latter sense since it is a convenient way to encompass all the processes occurring in a human body that maintain homeostasis.

Also, I will sometimes use the terms pathology and pathophysiology interchangeably. However, a distinction can be made so that these correspond to the study of abnormal conditions (pathology) and the specific processes occurring due to these conditions (pathophysiology). In this dissertation, I will take pathology to be the outcome of pathogenesis. This refers to a certain physical state of an organism identified with having a certain disease which give rise to specific symptoms and signs. This understanding of pathological states is in part motivated by Ereshefsky's (possibly eliminitavist) proposal that a certain physical state is designated as a disease by adding normative values to it, that is, whether we value or disvalue such a state (Ereshefsky 2009). For example, the pathogenic processes of erectile dysfunction are various. Erectile dysfunction can be caused by age related processes of losing cavernosal velocity and elasticity. Blood vessels can be damaged by insufficient blood sugar control due to type 2 diabetes. These processes result in a distinct pathological state or states identified as erectile dysfunction. The pathology of vascular erectile dysfunction may refer to the limited blood flow in the corpus cavernosum during erection due to the failure of smooth muscle cell

[2] https://www.webmd.com/a-to-z-guides/what-is-physiology

relaxation and vasodilation because of the low levels of synthesized nitric oxide (NO). But it can also refer to certain structural changes in the arterial and erectile tissue such as the increased wall-to-lumen ratio in the arteries (De Tejada et al. 2005). These pathological states ground the inability to achieve erection, and we disvalue the lack of erection for numerous reasons. Therefore, a rather straightforward view is that pathology is a description of some biological, chemical, or physical state of an organism that explains why and how some phenomenon occurs or does not occur. In Ereshefsky's view, one which I adopt here, pathology becomes a disease when we add values to these states.

To illustrate how all this comes together, observe how Rizzi and Pedersen (1992) define the "traditional" view of the causal chain of disease progress:

> (a) the disease manifestations (symptoms, signs, etc.), which are due to (b) the disease itself, which then again have their bearings from (c) a certain pathogenesis, all this banked on (d) an etiology.
>
> Rizzi and Pedersen 1992: 234

The idea that etiology and pathogenesis (and pathophysiology or pathology of a disease) do not involve the same causal processes and that they require different kinds of causal explanation is not a distinctive character of medicine and medical literature. Such claims are present in arguments by a number of philosophers of science (e.g., Salmon 1998, Glennan 2002, while Dammann in his 2020 book represents both disciplines, himself being an epidemiologist with a particular interest in philosophy of medicine and epidemiology).

But there is a difference between the views. Dammann's definition also includes pathogenesis and pathology as parts of etiology. In his view then, "the *etiological process* includes exposure and outcome as beginning and endpoint, as well as the pathogenetic mechanism(s) in between" (Damman 2020: 9). Such a view is perhaps held by the majority of medical scientists. For example, Dammann's view reflects the view stated in the well-known epidemiology textbook from McMahon and Pugh: "The etiology of a disease may be thought of as having a sequence consisting of two parts: (1) causal events occurring prior to some initial bodily response, and (2) mechanisms within the body leading from the initial response to the characteristic manifestations of the disease" (MacMahon and Pugh 1970: 26).

On the other hand, claims about different causal explanations coming from philosophy suggest a slightly different view among philosophers. For example, Wesley Salmon writes: "In many cases, I presume, causal explanations possess both etiological and constitutive aspects. To explain the destruction of Hiroshima by a nuclear bomb, we need to explain the nature of a chain reaction (constitutive aspect) and how the bomb was transported by airplane, dropped, and detonated (etiological aspect)" (Salmon 1998: 324). Following this distinction, a historian may give the informative etiological aspect of the explanation of the destruction of Hiroshima but the constitutive aspect – the nuclear fission – is a task for a physicist.

As I hinted above, I embrace the view that etiology refers to the factors or events which predate the onset of a disease while pathogenesis refers to the processes that lead from the onset of a disease to the manifestation of its symptoms and signs. So, to use Salmon's example, how the bomb ended up above Hiroshima is a distinct causal process, one which usually does not matter for the explanation of the causal processes constituting the chain reaction which occurred in the sky above Hiroshima. To translate this into a medical example, smoking two to three packs of cigarettes a day is in some cases an etiological cause of lung cancer. It explains how such a high level of carcinogens ended up in patient's lungs. On the other hand, the knowledge of how the destruction of cilia – a hair-like tissue that cleans the lungs – allows for these carcinogenic agents to accumulate and thus alter lung cells into cancer cells represents a step in the discovery of one of the pathogenic or pathological mechanisms of lung cancer. Whether a patient has contracted lung cancer by smoking two packs of cigarettes a day, a pipe, or rolled tobacco is not necessarily of use to scientists working in medical biological (basic) sciences when discovering molecular causal processes involved in the development of lung cancer or to medical practitioners who need to diagnose lung cancer and subsequently recommend the optimal treatment. Similarly, the claim that some proportion of lung cancer in a population is highly correlated with smoking habits in that population will be important for public health policy makers to predict the consequences for the prevalence of lung cancer if the smoking habits of cigarette smokers are lowered to some considerable degree (due to the implementation of a different public health policy).

In philosophy of medicine, a significant part of the discussion concerning causal explanations of medical phenomena have centered around the idea that etiology and pathogenesis are two different investigative strategies of causal inquiry in medicine. Each strategy has its own specific domain of interest, *disease causation* framework, causal concepts,

and epistemological and methodological frameworks. Representing those widespread sentiments, Damman argues that these two strategies require "two very different approaches that are provided by two scientific fields, epidemiology and basic science" (Dammann 2020: 9).

Nonetheless, this distinction is not a characteristic of contemporary Western medicine or philosophy of science. Differing views on the methodology of causal investigation in medicine have been around since at least the times of Hippocrates and Gallen. In the literature on the history of medicine these different approaches to disease causation are sometimes called *rationalistic* and *empiricist* approaches (e.g., in Newton 2001 or Bluhm and Borgenson 2011). This terminology, however, is not the same as the one between the different philosophical views of acquiring knowledge. In the medical literature, the distinction is between two often conflicting views of empirical inquiry of causal relations relevant for medical purposes and the type of evidence for causation that one finds compelling. The rationalistic approach emphasizes the importance of knowing the underlying physiological mechanisms that produce, influence, or constitute certain disease. It has also been referred to as the mechanistic or biological approach. On the other hand, the empiricist, difference-making, or simply epidemiological approach, in its modern version, is strictly a quantitative science. It uses statistical and mathematical methods to find correlations among variables and where causal relations are inferred from different individual cases or by comparing the differences in outcomes between populations. This approach is often presented (both as a praise and as a criticism) as a strategy interested only in whether something works regardless of knowing why or how it works.

In contemporary philosophy of medicine these two approaches are usually referred to as *difference-making* and *mechanistic* stances on disease causation. The predominant view in the philosophical literature is that counterfactual and/or probabilistic causal framework is the bedrock of the difference-making approach to causation. The correlations between values of variables are gathered either by observational or experimental studies where scientists observe and compare the outcomes of numerous similar cases. Clinical trials involving a potential new drug or vaccine present a good example. For example, we want to know if aspirin cures headaches. We give aspirin to patient *X* who has a headache. In some appropriate time, we observe and find out that the patient's headache vanishes (or that its intensity is significantly lowered). We could know whether aspirin relieves headaches if we could always go back in time and, in some considerable number of trials, have the same patient take or not take aspirin.

If in all or the majority of cases aspirin relieved the patient's headache, then we could conclude that aspirin cures headaches. But obviously, this cannot be achieved. So, we take a random sample of individuals from a population (taken to be significantly similar in their properties and indications) and provide them with aspirin. Even better, we split this sample of the population into two groups, one being administrated with aspirin and the other being administrated with a placebo, and then measure the headache-relief power of aspirin by comparing the outcomes in each group. Observe that the units of inquiry of the difference-making approach are populations. The conclusion that aspirin is the headache-relief factor is claimed when we have a considerable number of cases at our disposal. One headache relieved does not prove aspirin's causal difference-making status. Epidemiological research and its concepts, as will be developed in detail later, uses this difference-making account of causation.

Research in medical science has often begun after an observed regular association between two variables of interest, that is, between some factors, such as chemical compounds in the diet, specific physiological states, or behavioral patterns on the one hand and, on the other hand, specific health outcomes, such as cancer, heart attack, elevated blood pressure, etc. For example, John Snow's careful inspections of numerous households in the 1850s in London led him to pose a causal hypothesis connecting water pumps and the spreading of cholera. The rise of cigarette smoking and prevalence of lung cancer led to a famous study on the smoking habits of British doctors beginning in 1951, with the first reports about the observed correlation published in 1976 by Doll and Peto. Numerous observations of correlations between diseases and dietary, occupational, and similar factors were first made by life insurance companies; for example, the correlation between obesity and cardiovascular diseases (Hu 2008).

Observations of different factors that in some way influence health, lead to diseases, or are correlated with positive health outcomes were always present in medicine. In 19th century Europe, particularly in France, such an approach was called the numerical method, and had many advocates. Among these, Pierre-Charles Alexander Louis is probably the best-known. His work had a major influence on medical practice on both sides of the Atlantic. The numerical method advocated by Louis and some of his contemporaries consisted in collecting large amounts of clinical data, making group comparisons, and thinking in terms of populations rather than individuals (Morabia 2006). In other words, the numerical method exemplified the empiricist approach to disease causation in medicine. It can be seen as a precursor to modern epidemiological studies.

As mentioned previously, empiricist medicine was always a part of Western medicine, but its methods were fairly simple. Early epidemiological research (e.g., John Snow's work on the outbreak of cholera, Louis' group comparisons) lacked the sophisticated statistical tools and rigid methodology that are so characteristic of contemporary epidemiology. In this sense, the widespread appeal of empiricist medicine could not be achieved without a significant progress in statistics and probability theory. Although the rise of a contemporary version of empiricist medicine came swiftly, its epistemology and methodology developed slowly over the 20th century. Statistical and probabilistic sciences started to develop rapidly in the 20th century, but it took some time for them to enter medicine. Nonetheless, the numerical approach, group comparisons, and population thinking always lurked somewhere in the shadows of the medical mainstream and were advocated by a minority. Of these advocates, the most prominent names in the rise of epidemiology, clinical epidemiology, and the epidemiological approach in medicine in general were Ronald Fisher, A.B. Hill, and Archie Cochrane.

EPIDEMIOLOGY

What is epidemiology? Epidemiology is a scientific discipline full of conceptual, epistemological, and methodological issues which has only recently caught the attention of philosophers of science. What separates epidemiology from other fields of medicine? A frequently cited definition, which goes back to the 1970s and the aforementioned textbook by MacMahon and Pugh, defines epidemiology as the "study of the distribution and determinants of disease frequency in human populations" (Rothman, Greenland and Lash 2008: 32). It is a scientific discipline that lies at the intersection of several different sciences. Since it is concerned with the prevalence of disease and health outcomes in populations it is a major and important part of medicine – both theoretical and clinical. But epidemiology shares a lot of its methodology with sociology and economics. The units of inquiry of epidemiological studies, as can be seen in the definition above, are populations and the individuals composing populations.[3] Philosopher Alex Broadbent defines epidemiology in a very similar fashion but

[3] This does not necessarily mean that epidemiologists take populations as some kind of emergent entities, not reducible to mere aggregates of individuals. Although epidemiological research is concerned with populations, these can be understood as aggregates of individuals and the health of a population as the mean health of the individuals comprising the population. However, see Rose (1992) and (2001) for arguments in favor of the population stance rather than individual stance in epidemiological research.

he adds a methodological point into his definition. He defines epidemiology as the study of the distribution and determinants of diseases and health conditions of human populations by "means of group comparisons for the purpose of improving population health" (Broadbent 2013: 1). Most if not all epidemiological notions (e.g., risk ratios, odds ratios, relative risks etc.) are defined by comparisons between populations. In other words, epidemiological research does not aim to answer how a particular patient *X* developed high blood pressure but rather what are the factors correlated with chronic high blood pressure in some population under study. Epidemiological methods are then best described as methods of observation and classification or quantification and comparison.

Since the end of the Second World War, epidemiological studies have largely been concentrated on observing regularities and collecting statistical data in order to identify numerous *risk factors* found to correlate with certain health outcomes or diseases. That is, it comprises methods of observing how two variables of interest ″move together″ or of observing the ways that changes in the values of one variable are followed by changes in the value of another variable.

Figure 1. The black box approach to causation.

This empiricist approach – sometimes called *the black box* approach – does not claim that knowledge of the underlying biological or chemical mechanisms is unimportant. However, its proponents argue that epidemiological causal research should ignore that level of disease causation and only search for associations between variables. Therefore, correlations or associations "taken as ends in themselves" defined epidemiology as a scientific discipline from its very beginning (Kincaid 2011: 77). The black box approach, as some of its proponents have claimed (Peto 1984, Savitz 1994), proved to be very efficient in identifying cause-effect relationships in medicine (although epidemiologists have often avoided the explicit use of

causal terminology in their reports). Classical epidemiological methods of observational studies (cohort studies, case-control studies) found causal relationships decades before the underlying biological or chemical mechanisms were known (consider, for example, the cases of water pumps and cholera, smoking and lung cancer, or obesity and cardiovascular diseases). It is indisputable that we can know about a certain causal relation without knowing the exact mechanisms or processes underlying it. Furthermore, the black box proponents often claim that the urgency of medical situations allows us to ignore investigations into the underlying mechanism and ground medical interventions solely on the evidence of statistical associations.

MECHANISMS

The mechanistic stance, on the other hand, as I will use the term, stands for a completely different approach discerning causal relationships in medicine. It refers to research at the level of medicine's *basic sciences*: pathology, immunology, microbiology etc. By epidemiological population studies we have come to know that aspirin relieves headaches, but these studies remain silent on *how* aspirin does that – the black box remains black. What is it in aspirin and what exactly is the physiological target that makes aspirin an effective intervention? The population approach does not even aim at answering these kinds of questions. Therefore, the mechanistic stance or approach tries to open the black box and understand the processes, mechanisms, and causes by which a specific intervention or exposure factor is causally connected to the outcome. Hence, the mechanistic approach is often understood as "looking under the hood" approach. That is, it is concerned with explaining the correlations of the difference-making approach by discovering underlying causal processes that connect the administration of aspirin and the relief of headache. In that regard, it is often claimed that the mechanistic stance takes a rather different philosophical approach to causation and explanation, usually understood as a mechanical or productive account of causation: the activities and productive relations between numerous entities producing the effect or phenomenon are dependent on their properties and the features of their causal, spatial, and temporal organizations.

In the last 20 years mechanistic philosophy has gained much recognition in the philosophy of science. The rationalistic or mechanistic stance towards causal relations in medicine reflects the ideas of the modern-day mechanistic philosophers. It asserts that the

knowledge of causal mechanisms that are actually (and not plausibly or possibly) responsible for the phenomena of scientific interest provides epistemic grounds for achieving the different goals of medical science and practice, such as explanation or prediction (e.g., Machamer et al. 2000, Bechtel and Abrahamsen 2005, Craver 2007a), and making, evaluating, generalizing, and extrapolating causal hypotheses (e.g., Cartwright 2007, La Caze 2011, Clarke et al. 2013, Parkkinen et al. 2018). We should expect then that the inquiry into physiological, pathological, and pathophysiological mechanisms, and the subsequent use of the resulting knowledge for medical interventions and preventions is prominently involved in the satisfaction of different epistemic and action-oriented goals of medicine; for example, among others, claims involving the causes of diseases, pathophysiological mechanisms, claims about treatment procedures, assessments of efficacy and efficiency of drugs, and grounds for implementing health policies.

This is not just a trivial assertion. Rather, it has important epistemological consequences. It means that only knowledge of the productive steps of an actual mechanism connecting the exposure and outcome variables warrants a causal claim in medicine. Let me then refer to the evidence of underlying physiological, pathological, and pathophysiological mechanisms as the *evidence of mechanisms*. As mentioned, mechanisms in medicine should not only serve an explanatory purpose. We want to know what we can do with the knowledge of mechanisms in order to treat patients – we want to predict medical phenomena. Let us call this the *evidence from mechanisms*.[4] Having this second type of evidence implies that once the mechanism underlying the relation between exposure and disease (e.g., hypertension and heart failure) is known or established, it can (but does not need to) become evidence that a different causal relation will obtain under different circumstances – a certain counterfactual claim. It means that possessing a full (or partial) description of a mechanism establishes grounds for making claims about possible targets of interventions or predictions about the efficacy of these interventions. Possessing evidence from mechanisms means that we are in a position to infer from, for example, the evidence that high blood pressure causes different cardiovascular diseases to the claim that some specific treatment will stop or inhibit the effects of high blood pressure. In the philosophy of medicine inference from knowledge of mechanisms to means of interventions is often called *mechanistic reasoning* (e.g., Howick 2011a, Howick 2011b, Howick et al. 2010 Howick et al. 2013, Jerkert 2015, Solomon 2015). In the medical literature,

[4] I borrow this terminology from Jeffrey Aronson (2020).

however, such an inference is referred to as *pathophysiological rationale* (e.g., Guyatt et al. 1992, Montori and Guyatt 2008, Straus et al. 2018).

Discussions of contemporary mechanistic philosophy have predominantly been concerned with the explication of mechanistic explanation in molecular biology or neuroscience. Surprisingly, it is rather recently that mechanistic philosophers have turned their attention to mechanisms in medical sciences and the role of mechanistic knowledge in both theoretical and clinical medicine. Concerning the role of mechanisms in medicine, however, the discussion has primarily been concerned with the utility and scope of mechanistic knowledge of biomedical phenomena. Some of the often-discussed issues concerning the use of mechanistic knowledge in medicine include (i) the role of evidence of biological mechanisms in making causal claims of disease causation (e.g., Russo and Williamson 2007); (ii) the use of mechanistic knowledge in drug design, interpretation of experimental evidence, and extrapolation of evidence from experimental and observational studies (e.g., Steel 2008, La Caze 2011, Howick et al. 2013); (iii) predictions of outcomes of medical interventions based on the knowledge of underlying physiological, pathological, and pathophysiological mechanisms (e.g., Howick 2011a, Howick 2011b, Andersen 2012). Considering how important (iii) is in medical practice, there is no wonder that the discussion on this issue has been particularly vibrant and the most widespread in the literature. For example, clinicians working in intensive care units with patients who have coronavirus induced acute respiratory distress syndrome (ARDS) are faced with the question of outcomes of treating patients with ARDS by inhalation of nitric oxide (iNO). Similarly, researchers inquire about the mortality rate ratio between patients with ARDS who have been treated by iNO, and patients with ARDS who were treated by some other medical intervention. So, how useful is the knowledge of mechanisms involved in coronavirus induced ARDS in such cases? More importantly, can it be trusted? Can it yield accurate predictions?

This dissertation discusses both aspects of mechanistic knowledge in medicine (evidence of mechanisms and evidence from mechanisms). That is, I discuss what it means to give a mechanistic explanation of medical phenomena on one hand, and a prediction claim based on the knowledge of mechanisms on the other hand. Hence, each of the three chapters of this dissertation is concerned with a different aspect of the mechanistic stance in medicine. First, I discuss and define what makes the mechanistic approach different from the epidemiological or difference-making approach (from the epistemic, methodological, and

metaphysical point of view). In the second chapter, I discuss mechanistic philosophy's main or core claims, and provide my account of its metaphysical, epistemic, and methodological positions in the context of medical science and practice. Finally, in the last chapter I provide my analysis of mechanistic reasoning. I analyze reasons for its recurrent failure to give true prediction claims and discuss the structure of a good prediction claim based on the knowledge of mechanisms.

1. DISEASES, MECHANISMS, AND DIFFERENCE-MAKERS

Abstract

In this chapter I first present how the concept of disease causation has changed over time and how the focus of scientific research on different diseases has influenced our understanding of disease causation. Next, I discuss how the rationalistic and empiricist approaches are exemplified in the mechanistic and epidemiological approaches to diseases causation in contemporary medical science and philosophy of science. I analyze and discuss the main features of both approaches. I conclude that these approaches are concerned with causal explanations of processes occurring on different levels – the population level and the intra-individual level. Consequently, they require different kinds of evidence and evidence-gathering methods.

1.1. The multicausality of diseases in the early to mid-19th century

> Only in medicine are there causes that have hundreds of consequences or that can, on arbitrary occasions, remain entirely without effect. Only in medicine can the same effect flow from the most varied possible sources. One need only glance at the chapters on etiology in handbooks or monographs. For almost every disease, after a specific cause or the admission that such a cause is not yet known, one finds the same horde otf harmful influences—poor housing and clothing, liquor and sex, hunger and anxiety. This is just as scientific as if a physicist were to teach that bodies fall because boards or beams are removed, because ropes or cables break, or because of openings, and so forth.
>
> Henle, 1844; quoted in Carter, 2017, p. 24

In the passage quoted above, the famous German physician Jakob Henle expressed concerns which were shared by many of his contemporaries. Three features of mid-19th century medical science and practice can be extracted from this quote. First, since the mainstream view of the time asserted that numerous causes can equally contribute to the onset of any disease, reporting cases and making lists of all the known factors which have been observed to predate the occurrence of a specific disease constituted much of the work of medical science and practice. Second, Henle observes that there are no ontological constraints on what types or kinds of things are causes of diseases. The same disease can be caused by a certain behavior, drug, or food. Third, Henle claims that if medicine cannot explain why different things can cause a disease in one instance but in a different instance cause another disease, then medicine cannot be a science in the same way that physics is. That is, to be scientific, medicine needs a *theory of disease causation*.

At that time (the 18th century and early to mid-19th century), medical scientists and practitioners acknowledged several theories of diseases causation. The miasma theory (coming from the Greek word for pollution) proposed that diseases were caused by bad air – air polluted by emanations from rotting organic materials in the area. The humoral theory, on the other hand, identified diseases as imbalances between four humors in the human body: black bile, yellow bile, blood, and phlegm. To explain diseases, medical practitioners and scientists used concepts from one or the other theory. Nonetheless, it was not that uncommon to provide a causal explanation of some disease by incorporating elements from different theories of disease.

As mentioned, physicians often could not do anything else than make a report or give a list of purported causes for each disease. Ipso facto, the same cause could result in a different effect, that is, a different disease. Consequently, none of the reported and observed causes of some disease or state was thought of as being either necessary or sufficient cause of a disease. Since almost anything could cause anything, there could not be a consensus on how to achieve reliable or effective criteria for inferring causal relationships. So, for example, there were cases where a claim supported the miasma theory but not the humoral theory. Having a consensus on how to investigate causation by implementing causal criteria would have allowed medical hypotheses to undergo scientific scrutiny. This, to Henle's despair, was still absent in the pre-mid-19th century medicine. How, then, did physicians and scientists classify, characterize, and define diseases?

Causes of diseases, as scientists and medical practitioners in the 19th century usually claimed, were either *proximate* or *remote*. The proximate cause was primarily identified with a certain lesion or an "anatomical abnormality" (Carter 2017: 12). The primary source of knowledge of the proximate causes came from autopsies. Pathologists observed certain "morbid alterations" in a human body which they regarded as "the causes" of diseases (ibid). Remote causes "were factors normally external to the patient that explained the onset of disease" (Carter 2017: 13). They were further distinguished into predisposing or exciting causes. The knowledge of the remote causes was gathered mostly through observations and reports of individual patients or by patients' testimonies. The predisposing causes "were involved to cover characteristics of the individual's life or heredity that might render him or her unusually liable to a given disease" (Pelling 1997: 312). That is, predisposing causes were thought to make people more vulnerable to an exciting cause, which would trigger the onset of disease. Predisposing causes "render the body liable to become the prey of something, which has a tendency to excite the disease" (Elliotson, 1844, quoted in Carter 2017: 14). Sometimes, a predisposing cause was declared ineffective at bringing about a disease on its own. For the onset of the disease, it was presumed, there had to be an exciting factor which operated as a trigger. Exciting causes, therefore, were usually events that brought about some kind of mental or physical stress to a person (for example, wounds, inflammations, anxiety etc.). Predisposing causes involved a number of different behaviors, state of affairs, or properties, ranging from atmospheric pressure, geographical location, and climate to diet and unhealthy daily habits. Interestingly, both predisposing and exciting causes involved a lot of behavior that were seen

as deviations from the moral, religious, or social norms of the day. A lack of daily prayer was frequently identified as a predisposing cause.

As noted, since each disease could come about due to numerous different causes and each cause could result in numerous diseases, there was no consensus on the methodology for classification and definition of diseases other than by their symptoms. This is nicely depicted in the following example by K. Codell Carter: "hydrophobia was defined in terms of one prominent symptom: an extreme inability to swallow" (Carter 2017: 18). Considering the methodology of causal inference and the multiple theoretical frameworks of disease causation, diseases such as hydrophobia could not have been defined in a different manner. One physician could claim that the cause of hydrophobia is a physical alteration in the throat region while another could claim that it was due to some psychological trauma. The problem for 19th century medicine is that both physicians could be right. There was no point in undertaking research to associate any disease with a particular and universal cause. Carter sums this up nicely: "As long as diseases were defined in terms of symptoms, different episodes of any one disease simply did not share a common necessary cause. And no research, however brilliant, can find what isn't there" (Carter 2017: 36). Since the same disease could have been caused by a variety of completely different causes, it was an equally hard challenge to think of and implement successful and, more importantly, universal treatments. A treatment that would potentially work by eliminating one cause of a disease would be futile in treating the same disease when it had a completely different cause. Whether the hydrophobia of a particular patient was caused by a physical or psychological trauma would seem to be irrelevant. In that case, practitioners had no choice but to intervene into the symptoms of a disease rather than into its causes.

1.2. Infectious diseases, Koch's postulates, and the monocausal framework

Until the 20th century infectious diseases were the number one cause of death in both hemispheres. The impact of infectious diseases on human societies cannot be understated. The devastating consequences of infectious diseases were not reflected only in demographics; the economic and cultural consequences of epidemics such as bubonic plague or Spanish flu were immense. The consequences of several bubonic plague epidemics shaped the history of Europe and formed a lot of its folklore. By the mid-19th century diseases such as rabies, anthrax, cholera, and tuberculosis were still considered as the most dangerous threat to the overall health

status of human populations in every part of the world. Robert Koch, a famous German microbiologist, wrote in 1882: "Statistics show that one-seventh of all human beings die of tuberculosis, and that if one considers only the productive middle-age groups, tuberculosis carries away one-third and often more" (quoted in Carter 1987: 83). It should not be surprising then that infectious diseases were the focus of research in the medical and biological sciences of the 19th century. The interest in the causes of infectious diseases led to enormous breakthroughs in microbiology, technology, methodology of scientific research, and the formulation of new public health policies and their implementation.

By the beginning of the second half of the 19th century, the most important theoretical change in medicine was the acknowledgement of two general principles of the nature of disease causation by the medical mainstream: *the doctrine of specific etiology* and *the germ theory of disease*. The doctrine of specific etiology required that every disease have a specific cause rather than a whole collection or set of causes while the germ theory claimed that the causes of diseases are microscopic living organisms, alien to their host – germs. Therefore, the main goal of biomedical research was to show that for every specific infectious disease there was a specific and scientifically discernable microscopic causative agent. Famous figures such as Jakob Henle, Louis Pasteur, and Edward Klebs were some of the most influential scientists of that era. However, the history of medicine in the late 19th century cannot be written without mentioning the influential work of Robert Koch on the causes of anthrax and tuberculosis. By accepting both principles of disease causation Koch introduced to scientific medicine his postulates – often regarded as the first criteria for the assessment of disease causation.

Koch's postulates were intended as a series of methodological steps which, when implemented, would satisfy the goal of proving that a specific bacterium (e.g., *Bacillus anthracis* or *Mycobacterium tuberculosis*) is *the* cause of a specific disease (e.g., anthrax or tuberculosis). Interestingly enough, Koch never explicitly listed or named any of his postulates in the form of a definitive list. They are interpreted as such in the secondary literature. For this reason, the number of postulates usually ranges from three to five. Nevertheless, all interpretations agree on the core ideas of the postulates or their rationale. The postulates require the use of both observational and experimental method in a way very similar to the Mill's famous methods of causal inference (1843) – as both an observed and experimentally induced invariable regularity. The postulates demanded that in order to prove that a certain microscopic agent is indeed the cause of a disease (i) such an agent has to be observed in all cases of the

disease, (ii) it has to be isolated from the infected body and grown in pure culture in order to exclude any other potential biological material, and lastly (iii) it has to be transferred into another body in order to produce the same infection with the same symptoms. Much of the literature in philosophy of medicine (e.g., Broadbent 2009) and history of medicine (e.g., Carter 2017) asserts that Koch's postulates defined the cause of tuberculosis, and other infectious diseases, as being *universal* (according to the doctrine of specific etiology), *necessary*, since it is observed in all cases of the disease, and *sufficient*, but understood in a way that no case of the disease can arise without such a cause.[5] The research on infectious diseases influenced a conceptual shift in the classification of diseases. As a continuation of the scientific changes, that were initiated by the famous episode of Ignaz Semmelweis' work on the cause of the puerperal fever which changed the very definition of that disease, Koch's breakthrough work on tuberculosis changed the classification and definition of all infectious diseases. Vineis even calls the resulting conceptual framework the "Pasteur-Koch paradigm" because of its impact on the understanding of diseases, disease causation, and the methodology needed to discover their causes: "In the 'Pasteur-Koch paradigm' we find a clearly defined agent (usually a bacterium, a parasite or a virus) which is used as the 'unifying element' of a constellation of symptoms, i.e., the disease itself is largely defined and recognized on the basis of the agent" (Vineis 2004: 341). Henceforth, science started defining and classifying infectious diseases by their specific pathogens, which, in turn, were confirmed by Koch's postulates, and whose distribution and presence explained the pathogenesis of the disease (see Evans 1993, Carter 2017).

Figure 2. The monocausal model of disease causation.

[5] Ross and Woodward (2016), on the other hand, offer a different interpretation, along the lines of Woodward's interventionist theory of causation. I will present the main ideas of Woodward's interventionist theory of causation in section 1.4.2.

Semmelweis could only propose that the purported cause of puerperal fever as "the cadaveric material". However, by following Koch's postulates, scientists were able to specify the pathogen and describe its causal influence. Of course, Koch was well aware that to bring about tuberculosis in a specific individual, there would have to be a number of causes working together. But, as Broadbent (2009) points out, the criteria were intended to prove that only one of them had the properties of being both necessary and sufficient (given some general set of circumstances) to cause tuberculosis. Yes, an impaired immune system is just one of the conditions for the development of acute tuberculosis, however, the impaired immune system is not a sufficient condition to cause acute tuberculosis. Since there will be only one cause, which is both necessary and sufficient condition, such a disease causation framework is often called *the monocausal* model of disease causation. In some way, Koch's postulates are still in use today although they have been expanded and modified as criteria of disease causation to accommodate virus infections, parasitic infestations, and diseases of deficiency.

1.3. Back to multicausality: chronic non-communicable diseases

The discovery of antibiotics, vaccines, and other medical treatments significantly lowered the threat of infectious disease. In the 20th century the majority of population in the Western hemisphere experienced a profound change of lifestyle. This change, as is often emphasized, also resulted in a change of focus of medical science and practice. The attention of medical science moved to rather different types of diseases. These diseases were related to factors such as the sedentary lifestyle, changes in kinds of diet and dietary patterns, and other environmental and behavioral factors (such as the increase in smoking habits in the general population and stressful working and living environments). These kinds of diseases are not infectious and their influence on the health status of individuals and populations is not immediately noticeable. The onset of these diseases can take up to decades and, it seems, can be caused by a variety of different factors. These diseases include different kinds of cancer, diabetes, chronic lung disease, obesity, cardiovascular diseases, and hypertension – collectively called *chronic non-communicable diseases* (hereafter CNCDs).

As infectious diseases had accounted for the majority of deaths in the Western world at the time when Koch and his contemporaries were searching for their causes, in the second half of the 20th century the prevalence of CNCDs in Western population was at such a high level

that it was not unusual to view them as the modern-day epidemics. According to the WHO's *Global status report on noncommunicable diseases 2014*, in 2012 over two thirds of all deaths globally were due to a CNCD. Furthermore, contrary to a popular view, CNCD prevalence has decreased in high-income countries in the past two decades, while low-income and middle-income countries have experienced a significant increase. In the same report, it is stated that out of all global CNCD deaths, almost three quarters occurred in low- and middle-income countries. According to a WHO country profile (2018), out of 52000 total deaths in Croatia in 2016, nearly 92% were attributable to CNCDs: 45% of all deaths were due to some cardiovascular disease, 27% to some type of cancer, 4% were due to some respiratory disease, 4% to diabetes, and 13% to another CNCD. No wonder then that the focus of research in biomedicine, epidemiology, and clinical epidemiology had changed from infectious diseases to CNCDs since at least the 1950s.

The focus of scientific investigation of chronic disease causation led to theoretical and methodological changes. Recall that the acceptance of the doctrine of specific etiology and the germ theory of disease influenced scientists to think in terms of the monocausal model of disease causation: for every infectious disease there is a corresponding pathogen which is causally responsible for the onset of the disease. As stated in the previous section, a disease then became identifiable according to its specific pathogen. On the other hand, the monocausal framework of disease causation seemed not only a poorly suited framework to think about the etiology of CNCDs but quite possibly completely wrong.

To illustrate the above claim, consider the example of high blood pressure – hypertension. Normal blood pressure is of crucial importance to the healthy functioning of vital organs such as the heart or kidneys and the consequences of long-term hypertension are potentially devastating: e.g., stroke, myocardial infarction and heart failure, arrhythmias, and renal impairment. There are two different kinds of hypertension – primary and secondary – and they are differentiated by their different etiologies. Secondary hypertension is the condition of elevated blood pressure due to known causes, with the most common of these being tumors and some endocrine and kidney diseases. However, on average, secondary hypertension includes only 5% of patients suffering from elevated blood pressure. Primary or essential hypertension (hereafter EH), which makes up to 95% of all cases of hypertension, is defined as an elevated blood pressure which is due to unknown causes. This means that there is a collection of environmental and physiological factors (often working together) that, in some

way or another, cause or present a risk factor for the development of EH. Such factors include various unhealthy dietary behaviors, genetics, tobacco and alcohol intake, excess weight, or lack of physical activity.

Although the definition of EH is rather simple, the etiology of EH (and of all other CNCDs) is complex. The factors that figure in its etiology, just as much as the factors that constitute its pathology, are numerous and vastly different from each other. There is no equivalent of *Mycobacterium tuberculosis* for EH. For example, although tobacco and alcohol intake generally increase the probability of suffering from hypertension, they will not cause it in every individual. Similarly, many patients with coronary artery disease, heart failure, or survivors of heart attack suffer from EH through some substantial period of their lives but not all patients who had a heart attack suffered from EH nor do all who suffer from EH will have a heart attack at some time in their lives.

There is still some chance we will find necessary and sufficient causes for all cancers or for every instance of obesity and EH although as medical sciences progress this seems more and more unlikely. The majority if not all medical scientists agree that CNCDs do not have necessary and/or sufficient causes. The cause-effect relationships in all phenomena involving CNCDs are complex and multifactorial, and, perhaps, most appropriately expressed as counterfactual and probabilistic (Parascandola and Weed 2001). Not only has this led to different epistemologies and methodologies of causal inference in medicine, but it has also changed the definition and classification of diseases. Once again, the definitions of many CNCDs are now usually constitutive (identifying the constitutive mechanisms of the disease or their symptoms) rather than etiological.[6]

In her seminal and often quoted paper from (1994), Nancy Krieger argues that at least since the 1960s epidemiologists have used the notion of a "web of causation" to express this multi-causal framework where various causal pathways lead from factors of exposure to diseases and where different "chains of causation" intersect and share some of their steps. Krieger writes: "Expressly challenging the still-pervasive tendency of epidemiologists to think

[6] For example, consider obesity. In the latest WHO's International Classification of Diseases (ICD-11) it is "a chronic complex disease defined by excessive adiposity that can impair health. It is in most cases a multifactorial disease due to obesogenic environments, psycho-social factors and genetic variants. In a subgroup of patients, single major etiological factors can be identified (medications, diseases immobilization, iatrogenic procedures, monogenic disease/genetic syndrome)."

in terms of single 'agents' causing discrete diseases, the provocative metaphor and model of the 'web' invited epidemiologists to embrace a more sophisticated view of causality" (Krieger 1994: 890). The main contribution of the monocausal model of disease causation was a change to the way we think of what diseases are and how to classify them. It worked to some extent for infectious diseases and it can quite possibly still be a useful heuristic for thinking about diseases with bacterial or viral etiology.[7] But even in infectious diseases, not all patients who are diagnosed with an HIV infection will develop AIDS, nor will every patient who has contracted *Mycobacterium tuberculosis* develop acute tuberculosis. Nevertheless, diseases which are due to bacterial and viral infections always have their corresponding pathogen as a necessary condition. You cannot have tuberculosis without getting infected with *Mycobacterium tuberculosis* because tuberculosis is defined as a disease resulting from infection with *Mycobacterium tuberculosis*.

At least since John Stuart Mill's work in his *A System of Logic, Ratiocinative and Inductive* (1843) philosophical discussion on causation has been concerned with the idea that every effect will most likely have numerous causes. Mill observed that "[f]or every event there exist some combination of objects or events, some given concurrence of circumstances, positive and negative, the occurrence of which is always followed by that phenomenon" (Mill 1843: 237). Every case of causation will have background conditions that are indispensable for the occurrence of the effect. An effect *A* will normally have a whole cluster of factors, let us say *B*, *C*, *D*, *E* and *F*, that are essential for its occurrence as well as for the magnitude and timing of its occurrence. What we identify as *a* cause is just one factor in a whole set of factors necessary to bring about the effect. Infection with *Mycobacterium tuberculosis* is not by itself sufficient to cause tuberculosis. Other conditions, such as an absence of an immune response or antibiotic treatment, are also necessary for the onset of tuberculosis. *The* cause is the sum of all these factors that contribute to getting tuberculosis. The cause of *A*, then, is not *B*, or *C*, or *D*, but the conjunction of these factors, *BCDEF*. So, to be philosophically honest, there is no primacy to any of the necessary factors. Mill concludes: ″The cause, then, philosophically speaking, is the sum total of the conditions, positive and negative taken together, the whole of the contingencies of every description, which being realized, the consequent invariably follows″ (Mill 1843: 241). Only after all of the conditions are present, does the effect occur. It

[7] Think of COVID-19. It is a respiratory disease which shares multiple symptoms with other diseases, yet it is specifically caused by SARS-COV-2. Whether or not a patient with such symptoms has COVID-19 or some other disease is defined by the presence of infection by SARS-COV-2.

would not be misplaced to state that one should also have lungs to suffer from tuberculosis in the first place.[8]

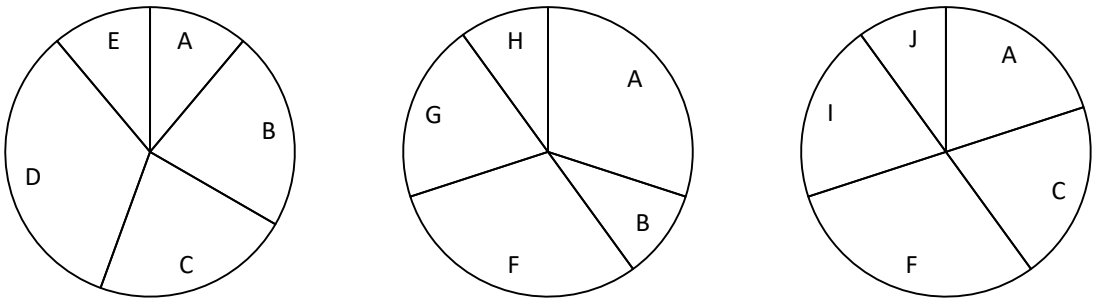
Following and expanding the ideas of Mill, the philosopher J.L. Mackie in his paper (1965) and later in a book (1974), and the epidemiologist Kenneth Rothman in his paper (1976) have come to very similar (almost identical) conclusions about the notion of cause. Consider Mackie's example of the fire caused by an electrical short-circuit (Mackie 1965: 245). The electrical short-circuit is only one of the conditions that are needed for the fire (others could be an inflammable material in the vicinity, presence of oxygen, absence of water sprinkles, etc.). As in Mill's account, the cause of a fire in a house, *P*, is a conjunction of these conditions *A*, *B* and *C*. However, Mackie notices that we can surely imagine numerous other conjunctions of conditions (for example, *DGH* or *JKL*) in which the fire could have started: "It may well be that *P* occurs only when at least one of these conjunctions has occurred soon before in the right region. If so, all P are preceded by (*ABC* or *DGH* or *JKL*)" (Mackie 1980: 61). If the total cause of the fire is this complete disjunction of conjunctions, in what way, then, do we (or the firemen) acknowledge the electrical short-circuit as *the* cause of the fire? Mackie says that the conjunction of conditions *ABC* was sufficient for the fire, but it was not necessary since the fire could have started in some other way.

How does this apply to medicine? Causes of some CNCDs like EH will be numerous and different. It seems a rather pointless attempt to list the whole conjunction of conditions which are necessary for the consequent – EH – to invariably follow. Not all smokers will have high blood pressure, yet many cases of high blood pressure can be linked with tobacco smoking. There will be one set of conditions which will be sufficient to bring about the condition of high blood pressure (e.g., *ABC*) yet, contrary to Mill, this set will not be necessary since there will be another set equally sufficient (e.g., *CEFG*) but which will lack some or all of the conditions present in the first set. Therefore, both smoking and genetic predispositions can be considered as *a* cause of EH but are neither necessary nor sufficient for it. In cases where smoking positively contributed to EH, smoking will be one of the factors in the whole set of

[8] This leads to the problem of causal selection: why are we ready to accept that *M. tuberculosis* is the cause of tuberculosis, but the presence of lungs is not? Similarly, why do we not recognize the presence of oxygen as a cause of the lighting of the match? The problem of causal selection is an interesting one and there is an enormous literature on it in philosophy and different scientific fields. For the clarity of the dissertation, I will not engage in this discussion further. It should be noted, however, that philosophers and philosophically-minded scientists are aware of the causal selection problem.

conditions which was sufficient to bring about this case of EH. What, then, is the causal status of the electrical short-circuit in the case of a fire, or smoking in the onset of EH? For Mackie, the full cause of the disease *D* is the disjunction of conjunctions (*ABC* or *CEFG* or *JKL* or …) and what we ordinarily consider a cause is actually only an *INUS* condition that is, an **I**nsufficient but **N**ecessary part of an **U**nnecessary but **S**ufficient condition.

Completely unaware of Mackie's work (to his confession), epidemiologist Kenneth Rothman developed his version of the INUS framework to think about causes in medicine. In addition to being a metaphysical analysis, his discussion on the notion of cause was intended to have practical consequences too. Similar to Mackie and Mill's claims about causation in the sciences, Rothman claims that medicine has used the term cause only for specific conditions which are insufficient to bring about a disease themselves. In the case of EH, a lack of exercise is not by itself sufficient to cause high blood pressure. Nevertheless, when a lack of exercise is combined with other conditions which are known to present serious risk in developing EH, such as high intake of salty food and smoking tobacco, their full set can become sufficient for EH. Furthermore, there can be several distinct sets of conditions which are together sufficient (see Figure 3). This approach is useful for the health sciences since such sufficient sets have the "restriction to the minimum number of required component causes; this implies that the lack of any component cause renders the remaining component causes insufficient" (Rothman 1976: 591). This offers a rationale for medical interventions: an intervention can be performed on just one of the component causes since it is presumed that removing or preventing the action of one of the components will prevent the onset of the disease, at least through one of its possible multiple causal mechanisms.



A – unknown causes; B – smoking; C – genetics; D – lack of exercise; E – salty food; F – excess weight; G – chronic kidney disease; H – stress; I – diabetes; J – old age.

Figure 3. Rothman's causal pies for the case of hypertension.

Nonetheless, Rothman's account says little about the methodology used to find or discover each component.[9] Also, there might be case where we know a great number of factors of some sufficient-condition constellation, yet where the effect does not occur. The sufficient condition account then relies on the supposition that there are no hidden, unknown factors that exert too much of an influence in the case. To account for the latter possibility, there has to be a probabilistic reading of the causal influence of every component even though the constellation of conditions, when completed, is a deterministic (i.e., sufficient) cause. Next, sufficient-component models or pie-charts do not say anything about the time-sequence (a

[9] The most prominent frameworks for these particular types of questions and studies include, for example, the potential outcomes approach and structural equations models. The notion of cause that such frameworks use is counterfactual and/or probabilistic. The potential outcomes approach (e.g., Rubin 1974, Holland 1986, Hernan and Robins 2020) is close to the counterfactual understanding of causation in philosophy (Lewis 1973, Woodward 2002, 2003). The literature on the apparatus for probabilistic causation used in social sciences and epidemiology is immense (e.g., Pearl 2000). For example, observe two definitions of causation in the epidemiological literature: "A causal association is one in which a change in the frequency or quality of an exposure or characteristic results in a corresponding change in the frequency of the disease or outcome of interest (Hennekens and Buring 1987:30)", and "a cause of a disease occurrence is an event, condition, or characteristic that preceded the disease onset and that, *had the event, condition, or characteristic been different in a specified way, the disease either would not have occurred at all or would not have occurred until some later time*" (Rothman, Greenland and Lash 2008: 6, emphasis added). The potential outcomes approach relies heavily on randomized experiments which are now set as standards for causal inference in medicine, but observational studies also follow its rationale. I will address randomized experiments and other epidemiological studies as well as their philosophical background in section 1.4.2.

component A can exert its influence for 20 years while B only for 5 years or months), or the dose of the component and its contrast (for example, smoking 20 cigarettes a day rather than 10 cigarettes a day). The model, therefore, has to be extended to include a description of contrast of index and reference conditions that defines each component cause (see Rothman et al. 2008). In spite of these inadequacies, sufficient-condition model of disease causation helps us to imagine what the links in Krieger's web of causation stand for. In that regard, it has proven to be a useful heuristic for depicting how different sets of conditions participate in the etiology of a given disease by combining into various causal mechanisms.

An important consequence of this account in terms of doing research is that it does not impose any ontological restriction on component causes. HIV virus will be a necessary component in all of the disjuncts or constellations for HIV infection. However, in designing constellations of conditions we can make a constellation where some of the disjuncts will be types of behavior or one-off events such as unprotected sexual intercourse. On the other occasions, disjuncts might include conditions such as the transfusion of infected blood. Etiological causal explanations of HIV infection or EH in Rothman's account evoke causes which occupy different levels of a causal network (in Krieger's terms) – from the molecular mechanisms of viral replication after the HIV virion enters the body to frequent unprotected sexual intercourses or the lack of appropriate screening of blood donors. Although I discuss these issues in much more detail in the next section, it is worthwhile to mention a point here to anticipate the reader's possible concerns. Causal pathways in the web of causation for EH can include different food patterns or lack of exercise (when these variables are properly defined so that we can have means of their measurement – cf. Holland 1986) as well as different molecular or genetic pathways. Both types of causal information are important, and neither can replace the other. For example, it is important to know the molecular pathways by which HIV attacks T-cells, but this type of information will not be useful to explain why the prevalence of HIV infections in some region is on the rise over the last 10 years. The rise of HIV infections might be explained more successfully for example, by lack of donor screening which, I presume, would not be an explanation that cites molecular pathways.

To conclude, Rothman's model of disease causation fits well in the multiple fields and studies of medicine - knowledge about all the components that figure in some sufficient set of conditions is incorporated from various studies: from statistical epidemiological studies to the findings of basic medical sciences such as pathology, pathophysiology, or immunology. Here,

it served the role of presenting how causal claims found in different fields of medicine use different causal notions and causal concepts. Equally importantly, causal explanations of diseases involve different causal relations on different levels – they talk about biochemical pathways and physical traumas, and they talk about frequent unprotected sex or smoking. They use both populational talk in discussing diseases and disease causes, and they talk about particular events and their particular causes. Yet they all talk about the same thing – the causes of diseases or the means and outcomes of treatments.

1.4. Two approaches to disease causation

Empiricist and rationalistic medicine represent the two dominant perspectives or approaches to disease causation in Western medicine. Though their metaphysical, epistemological, and methodological commitments have changed significantly over times, the enduring characteristic of the debate is a difference in their views on the usefulness of the knowledge of underlying biological causes. The rationalistic or biological approach and its metaphysical, epistemological, and methodological commitments dominated both the theoretical and practical aspects of Western medicine for most of the 20th century. Mechanistic philosophy, on the other hand, experienced its revival only at the end of the century. Right at the time when mechanistic philosophy started to enter the philosophical mainstream, a new version of medical empiricism was on the rise.

In this and the following section I will discuss what these two approaches to disease causation in modern medicine (at least since the second half of the last century) amount to. They both aim to explain diseases and to discover causal relations leading to diseases or constituting the pathology of a disease. Nevertheless, as will be argued later in the text, these two frameworks aim at explaining diseases on different levels: the intra-individual versus the population level.

Some medical scientists and practitioners would strongly disagree with my equation of rationalism in medicine with biomedicine but henceforth I will use these terms interchangeably. Why would such a usage of these terms be controversial or at least disputed by some? By the term biomedicine some authors refer to contemporary Western medicine in general. These authors contrast biomedicine with general medical frameworks of the past (with their own metaphysical, epistemological, and methodological commitments) or with others that are

supposed to be incommensurable to the science and practice of contemporary medicine in the Western world. Examples of such medical frameworks would be Hippocrates' medicine or traditional Chinese medicine (for example, consider Sean Valles' entry *Philosophy of Biomedicine* in *The Stanford Encyclopedia of Philosophy*, Shapiro 2003, and Clarke and Russo 2018). Considering the difference between rationalism and empiricism in medicine, then, the discussion would be strictly connected to the epistemology of causation in medicine. In that regard, both frameworks are just different parts or aspects of biomedicine in general. However, I will refer to rationalism in medicine as the biological, biomedical, or mechanistic approach and empiricism in medicine as the epidemiological or statistical approach (that is, the Evidence-Based Medicine framework, but more on that term later) for reasons to be discussed below.

1.4.1. The biological or mechanistic approach

One of the key events in the rise of the biological approach in medicine was the so-called Flexner report to the Carnegie Foundation by Abraham Flexner in 1910. Flexner, a schoolteacher and an educational expert, toured Europe, Great Britain, and the United States, and visited their medical schools.[10] He was invited by the Carnegie Foundation to write a report on medical education in the United States (Duffy 2011). In it, Flexner advocated for a scientific approach to medical education where scientific medicine was understood as mechanistic science grounded in laboratory work. Medical education, according to Flexner, should consist of medical theory (evidence gathered from the laboratory by in vivo and in vitro experiments) and clinical practice where the theoretical medical knowledge would finally be tested and used. This Flexnerian medical education influenced and changed the way medicine was taught on both sides of the Atlantic. It brought structure to medical education to which most of today's medical schools still adhere to. Until the late 1990s practicing clinicians all over the world were trained to ground their inferences in the diagnosis, prediction, and treatment of individual patients on mechanistic biomedical knowledge (anatomy, physiology, pathophysiology, immunology etc.) combined with their bedside experience.

[10] Interestingly, Abraham Flexner majored at Johns Hopkins University in Greek and Latin and philosophy.

At least since Flexner's report, the biomedical approach, exemplified by laboratory sciences and the biological conception of disease and health, has been the predominant view of disease, medicine, and medical education in the Western world. As such, it constitutes a philosophical, theoretical, and methodological framework within which scientists define and investigate human health and diseases and develop treatments. It sees medicine as an extension of biology in practice, that is, as "nothing but applied biology".[11] Its first major breakaway happened in the late 19th and early 20th century with the development of laboratory techniques and methodologies and subsequent important medical discoveries such as the isolation of insulin. It started with the germ theory, the doctrine of specific etiology, and the incredible breakthroughs of biological sciences in the late 19th century, and continues all the way to this day with, for example, precision medicine and genetic treatments.

Whether the biomedical framework constitutes a paradigm of medicine in the full Kuhnian sense is definitely open to discussion. It is certainly conditioned by what we mean by biomedicine. As mentioned previously, if biomedicine is Western medicine in general, with both of its causal frameworks (biomedical, laboratory sciences and epidemiology), and if its rival paradigms are non-Western medical frameworks, for example traditional Chinese medicine, then, biomedical concepts do seem incommensurable with those of different frameworks (paradigms). However, I am referring to biomedicine as a medical practice that is exemplified by laboratory sciences, their methodology of causal inference, and biological conceptions of disease, disease causation, and health. In that case my term biomedicine refers to what is usually called medicine's *basic sciences*. Still, however, I believe we can identify certain biomedical metaphysical, epistemological, and methodological commitments which constitute the biomedical framework or approach – its view on health and disease, causes of diseases, and the methodological, epistemological, and evidential frameworks it uses to investigate and explain diseases and their causes. Nonetheless, for reasons I will explain later, I do not think that these are incommensurable with the empiricist approach to medical practice.

The metaphysical, epistemological, and methodological commitments of biomedicine are nicely described by Marcum:

[11] Valles, S. (2020). Philosophy of Biomedicine. *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2020/entries/biomedicine/.

> Starkly put, the patient is reduced to a physical body composed of separate components that occupy a machine-like structure. The biomedical practitioner's emotionally detached concern is to identify a patient's diseased body part(s) and to treat or replace the diseased part(s) in a fashion analogous to a mechanic. The outcome of this intervention—curing the patient— derives from specifically diagnosing a diseased or dysfunctional part(s) and then treating, scientifically, the cause of the disorder. The results are commendable: to cure disease, to relieve pain, and to prevent death.
>
> Marcum 2008: 393

The biomedical model (or as it sometimes called, "the old medical model" (Fuller 2017)) is marked by a so-called ″biological chauvinism″ (Broadbent 2009). That is, biomedicine's metaphysical commitments are (i) naturalistic in the sense that is characteristic of biological conceptions of life, health, and disease (against vitalism), and (ii) reductionist in the sense that an explanation of a medical phenomenon is explained in terms of its parts. These commitments further include three distinct epistemological and methodological theses of biomedicine. First, the conception of the human body as decomposable into biological parts and biochemical processes and pathways that maintain the normal functioning of the body. Second, biomedicine searches only for an organic, objective entity as a cause of disease (Marcum 2008). That is, what diseases and their causes are "is restricted to solely biological, chemical, and physical phenomena" (Krieger 2011: 130). This concept or model of body functioning has its roots in the work of the famous French physician Claude Bernard and the idea of *homeostasis* as "equilibrium within the body despite changes in the internal or external environment" (Lakhani et al 2009:3). This leads us to the third thesis. Diseases occur when the normal functioning of bodily mechanisms is in some way disrupted (the most well-known naturalistic theory of health and disease is Boorse's in his (1977) and (2001)). Knowing the mechanisms and processes of homeostasis and how their disruptions or dysfunctions lead to diseases served as the basis for the design of medicinal interventions for most of the 20th century.

Characteristically mechanistic or biomedical medicine is represented by medicine's basic or bench sciences. These are biological sciences that study all the biological causes, mechanisms, and processes within the human body - pathology, physiology, microbiology, immunology, or pharmacology. Biomedicine's basic or bench sciences are laboratory sciences. They study phenomena by performing in vivo and in vitro experiments (however, in silico

experiments are also regularly used in biomedicine too). These experiments are performed on animals models (for example, laboratory mice or fruit flies). The domain of phenomena that these sciences study is biological and chemical. Their explanations are characteristically mechanistic and sometimes mechanical (as in the various models of mechanisms of injury, specifically with broken tendons, bones, and joints). I will present and discuss these notions in full detail in the next chapter. For now, I will only note that the characteristics of explanations in biomedicine are similar to Wesley Salmon's theory of causation and explanation (e.g., in his (1998)) and mechanistic theories of causation and explanation (as in Glennan 1996, 2002, or Machamer, Darden and Craver 2000). The focus of those sciences is on the cellular, metabolic, and other processes occurring in the body and how different diseases influence these processes in ways we consider to be pathological. Lakhani et al. start their textbook on basic pathology with the following passage: "This book will adopt a strongly *biomedical concept* of disease. This is a *mechanistic model* that regards the body as a machine with repairable or replaceable parts. It looks for specific *underlying biological causes* and places a high emphasis on the scientific evidence-base for untangling cause and effect in both the disease and its treatment, because this is important for patient care and prognosis" (Lakhani et al. 2009: 3, emphasis added). The rationale is rather simple. Knowing how particular biological mechanisms "normally" function is required for knowing all the different ways in which they can be disrupted or impaired.

Of course, the concepts of "function" and "normality" of biological mechanisms are highly controversial in philosophy. I will leave the discussion on these issues for the next chapter. However, for the present purposes it is worthwhile to mention Boorse's definition of normality in his (1977) since it is used in his definition of health and disease, a definition which is often connected to the metaphysical commitments of the biomedical approach.

When diseased, we feel something is wrong with our body or that some bodily functions are not working "properly". Disease means that our body or parts of our body are in some kind of a state which, we presume, is non-natural or it is out of the ordinary. As Boorse says: "Health is functional normality, and as such is desirable exactly insofar as it promotes goals one can justify on independent grounds" (Boorse 1975: 60, 61). What are the characteristic features of diseases, that which make a part of our body or our body as a whole to be in a non-natural state or at least make it undesirable for us? Disease is undesirable because it impairs survival and reproduction of an organism. Boorse takes these to be "independent grounds" which make his

theory natural and not created, influenced, or shaped by cultural and social factors. However, both aspects of his theory are clearly controversial. Not all disease states impair our survival and reproduction, while some states that do so are not considered diseases. On the other hand, defining normality or normal function is a notorious area of trench warfare in the philosophy of science. So how does Boorse answer these worries?

As noted, Boorse intends his theory to be a value-free theory (free of any cultural and social influence). Boorse, then, turns to physiological and pathological knowledge together with evolutionary biology to construct an objective theory of health and disease. Since this account defines disease as deviation from a physiological statistical norm, it is usually called a *Biostastical theory* (BST). So, Boorse defines normality of a "part or process within members of the reference class as a statistically typical contribution by it to their individual survival and reproduction" and where the reference class is a "natural class of organisms of uniform functional design", that is, an age group of a sex or species (Boorse 1977: 562). Therefore, "[*h*]*ealth* in a member of the reference class is *normal functional ability*: the readiness of each internal part to perform all its normal functions on typical occasions with at least typical efficiency", and a "*disease* is a type of internal state which impairs health, *i.e.*, reduces one or more functional abilities below typical efficiency" (ibid.).

Obviously, the notions of reference class and normal functioning have been highly disputed since Boorse first proposed BST (see Kingma 2007). There is a sense in our understanding of health and disease that includes some kind of value evaluative component. That is, not everything about our conception of disease is explained in biological terms. Theories that incorporate the notion of value in their definition of health and disease have been proposed as an answer to Boorse's BST – collectively, they are usually called *normativist* theories (for example Engelhardt 1986, or Wakefield 1992).[12]

[12] These theories avoid metaphysically loaded notions such as normality and function. They especially focus on what we cherish and desire as human beings. For example, notice how Engelhardt argues that health "must involve judgments as to what members of that species *should be able to do*—that is, must involve our esteeming a particular type of function" (Engelhardt, 1976: 266, emphasis added). By putting values at the center, normativists argued that they can explain why some cultures think of some states or conditions as diseases while others do not. In addition to adding our values to the definitions of health and disease, normativist theories incorporate things that Boorse's theory excludes – mental diseases and various non-physiological causes of health and disease, such as a stressful environment. However, there are many examples that present problems for normativist theories. There are states that are considered as undesirable or unpleasant, but it is certainly controversial to designate them as diseases – e.g., PMS, alcoholism, being overweight or obese (Ereshefsky 2009). Furthermore,

Biomedical or mechanistic medicine, therefore, is the way of "scientific medicine", so to speak. It is implied that identifying the causes of diseases from the biological, chemical, or physical domain and offering a causal explanation of the disease phenomenon in terms of their causal properties is "more scientific" than the strictly pragmatist approach of empiricist medicine (characterized by statistical, probabilistic methods). More importantly, as many have argued, if we know the mechanism of a disease, then, not only can we explain the association between a putative cause and a disease, but we can also infer further claims about the outcomes of interventions. For example, observe this passage by Claude Bernard in his *An Introduction to the Study of Experimental Medicine*:

> Now that the cause of the itch is known and experimentally determined, it has all become scientific, and empiricism has disappeared. We know the tick, and by it we explain the transmission of the itch, the skin changes and the cure, which is only the tick's death through appropriate application of toxic agents. No further hypotheses need now be made about the metastasis of the itch, no further statistics collected about its treatment. We cure it *always* without any exception, when we place ourselves in the known experimental conditions for reaching this goal.
>
> Bernard 1999: 214

Such biomedical commitments and promises are shared by modern day mechanistic philosophers. Machamer, Darden and Craver in their (2000) express this clearly. They claim that mechanisms or their representations in scientific research "are used to describe, predict, and explain phenomena, to design experiment, and to interpret experimental results" (Machamer, Darden and Craver 2000: 17). Both quotes express the same view. The knowledge of biological causes or mechanisms has two roles. The first one is the explanatory role – it is the *evidence of mechanisms*.[13] This is what Machamer, Darden and Craver mean when they say that mechanisms are used to describe and explain. It is the knowledge of the underlying physiological mechanisms supporting and maintaining homeostasis, and the pathological and pathophysiological mechanisms issuing at the onset and development of diseases and manifestations of their symptoms and signs. In a nutshell, evidence of mechanisms is any

normativist theories seem to justify the treatments of homosexuals in the past because, at the time, a particular society held it as an undesirable state or condition which required a medical treatment.

[13] Recall from the *Introduction* that the evidence of mechanisms is the knowledge about the physiological, pathological, and pathophysiological mechanisms underlying some medical phenomenon.

physiological explanation of a biomedical phenomenon. But these authors also claim that mechanisms in medicine have a further role. We want to know what we can do with the knowledge of mechanisms in order to treat patients. That is, mechanisms ought to have a role in devising medical intervention procedures and in making predictions about the efficacy of these interventions – that is, a predictive role.

The arguments from Bernard and modern mechanistic philosophers for this second type of evidence concerning biological causes and mechanisms state that once the mechanism underlying the relation between exposure and disease (for example, hypertension and heart failure) has been established, it can become a piece of evidence that a different causal relation will obtain under different circumstances. In other words, knowing the biological causes of bodily functions and disease pathogenesis should allow making true counterfactual claims. This means that possessing a full description of a mechanism allows making claims about interventions (as means of treatment) and predictions about the effectiveness of those interventions. As noted above, inferences drawn from the knowledge of mechanisms to means of interventions is often called mechanistic reasoning in philosophical discussion while in the medical literature it is referred to as a pathophysiological rationale. Howick defines mechanistic reasoning as involving "an inferential chain linking the intervention (such as antiarrhythmic drugs) with a clinical outcome (such as mortality)" (Howick, Glasziou and Aronson: 2010: 434). This is the type of evidence that tell us that if we know how essential hypertension can cause various cardiovascular diseases then we should also be in a position to know how to develop specific treatments, for example, by developing different drugs which target some of the steps or entities in this mechanism of disease.

To conclude, the biological-mechanistic approach is a framework for the investigation of disease causation and biological functions in the human body based on laboratory sciences and their methodology. It aims at revealing pathogenic, pathophysiological, and physiological processes in the human body. The knowledge gathered through laboratory experiments ought to provide grounds for two different types of claims about the presence of causal relations involving human health and disease. The first is concerned with the functioning of physiological processes maintaining health and leading to diseases. The other is about making predictions of the outcomes of medical interventions. Medical interventions are rarely if ever made without the knowledge of physiological, pathogenic, and pathological mechanisms but predicting their efficacy, as I shall discuss later, is a different matter.

1.4.2. The epidemiological or difference-making approach

In the late 1980s and early 1990s, a group of epidemiologists, primarily centered at Oxford and McMaster University in Canada, started to question biomedical epistemological capabilities to produce true prediction claims relevant for the treatment and prevention of diseases. That is, they argued for a change in approach to causal reasoning. Although the criticism was primarily aimed at biomedical epistemological capabilities, it eventually led to the development of an all-encompassing theoretical, methodological, and evidential framework for thinking about health and disease, discovering disease causation, and evaluating the efficacy of medical treatments. Its influence on medicine was so great that some even declared it as a "new medical model" (Fuller 2017). Although the first paper introducing the novel concept "Evidence-Based Medicine" was published in 1991 by Gordon Guyatt, the publishing of the "EBM manifesto" in 1992 in the *Journal of the American Medical Association* represents its starting point in earnest. The individuals from McMaster University, formed the Evidence-Based Medicine Working Group (consisting of, among others, David Sackett, Gordon Guyatt, Brian Haynes, and David Churchill), and published the "EBM manifesto" where they confidently announced that a "new paradigm for medical practice is emerging" (1992: 2420).[14]

Bluhm and Borgerson (2011) observe that three aspects of basic science's influence on clinical practice were influential in the rise of EBM movement: "the growth in laboratory research in medicine, the growth in clinical research in medicine, and the realization that, despite the increase in scientific knowledge, medical practice was not uniformly influenced by the results of research" (2011: 206, 207). The progress of biomedical sciences throughout the 20th century was immense. We have learned a lot about the human body but seldom did this knowledge translate into clinical practice. On the other hand, clinical research changed completely with the development of randomized controlled trials. First used in agriculture, they are now mostly associated with medicine. Following the Second World War the randomized controlled trials technique was adopted to test interventions in medicine, and the first such trial

[14] "What on earth has medicine been based on before?" Worrall's imaginary newcomer amusingly asks of EBM at the beginning of his paper (2002: 316). Indeed, to a newcomer the "evidence" part of the name of this new approach might sound confusing since it implies that medicine was not based on any kind of evidence before the rise of EBM. Solomon argues that "evidence" in EBM was meant to have "rhetorical power, because who would deny the importance of evidence?" (Solomon 2016: 289). So, it is not controversial to claim that perhaps EBM is a kind of misnomer. For these reasons Solomon argues that EBM is perhaps better called "epidemiological medicine" (Solomon 2016: 289) since, as will be shown later in the text, it most values evidence which has been gathered by epidemiological methods.

used in medical context is often credited to Austin Bradford Hill and his work on streptomycin as a treatment of tuberculosis in 1947 (Solomon 2011, Clarke et al 2013). The epistemological superiority of randomized controlled trials was praised immediately, and already in the late 1970s and early 1980s they were recognized as "the golden standard" for assessing the efficacy of medical interventions. As David Sackett, one of the most prominent figures in the development of the EBM movement, argues, in contrast to laboratory work and findings from basic sciences, the results acquired by clinical epidemiological studies are "immediately applicable" (Sackett 2000: 380). Recall that it took decades before we had a description of mechanisms that lead from carcinogens in cigarette smoke to lung cancer. The evidence gathered through trials, on the other hand, was praised exactly for its lack of dependency on biological theory. No need to wait for the laboratory; EBM's "protocols" were intended to be applicable immediately to clinical practice through the implementation of the up-to-date statistical evidence.

However, it is worth pointing out that EBM was not only a response to problems with translating biomedical knowledge into clinical practice. It was equally a response to the so-called "authority of expertise" in clinical practice. There were two aspects of the authority of expertise which the EBM-ers were eager to argue against. The first aspect which EBM-ers tried to minimize was reasoning based on experience gathered "by the bedside". Practitioners based much of their faith in medical interventions on their previous positive or negative clinical experience combined with their biomedical knowledge. The second aspect of the authority of expertise was in how this figured in developing medical guidelines.[15] Prior to EBM's criticisms, guidelines were developed by authorities in the field based on their experiences. EBM, as we shall see, was imagined as a democratic and progressive new approach that tried to minimize the authority of expertise. Firsthand experience and the authority of expertise was labelled as lacking in scientific rigor and testability. Expert opinion and consensus conferences were seen as the way of the old where men financed by pharmaceutical companies gathered around the table to figure out the guidelines which should then be imposed on practitioners in their everyday clinical work. An interesting EBM perspective on the matter is given by Trisha Greenhalgh:

[15] Guidelines are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" (Institute of Medicine 1990).

> When I wrote the first edition of this book in the mid-1990s, the most common sort of guideline was what was known as a *consensus statement* – the fruits of a weekend's hard work by a dozen or so eminent experts who had been shut in a luxury hotel, usually at the expense of a drug company. Such 'GOBSAT (good old boys sat around a table) guidelines' often fell out of the medical freebies (free medical journals and other 'information sheets' sponsored directly or indirectly by the pharmaceutical industry) as pocket-sized booklets replete with potted recommendations and at-a-glance management guides. But who says the advice given in a set of guidelines, a punchy editorial or an amply referenced overview is correct?
>
> Greenhalgh 2019: 7

Authority of experience and expertise was seen as biased, undemocratic and, most importantly, fallible, and therefore unreliable. To achieve the high standard that they advocated, EBM-ers argued for a new approach to evidence in medicine. This new approach argued for a systematic, unbiased, and democratic approach to developing guidelines and evaluating causal claims relevant for clinical practice. It was supposed to "de-emphasize intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making" (Evidence-Based Medicine Working Group 1992: 2420) and use only the "current *best evidence* in making decisions about the care of individual patients" (Sackett et al. 1996. BMJ. 312: 71, emphasis added). An all-encompassing definition of the goal that EBM tries to achieve and how this should be achieved is probably best given in Davidoff et al (1995):

> In essence, evidence based medicine is rooted in five linked ideas: firstly, clinical decisions should be based on the best available evidence; secondly, the clinical problem – rather than habits or protocols – should determine the type of evidence to be sought; thirdly, identifying the best evidence means using epidemiological and biostatistical ways of thinking; fourthly, conclusions derived from identifying and critically appraising evidence are useful only if put into action in managing patients or making health care decisions; and finally, performance should be constantly evaluated.
>
> Davidoff et al. 1995: 1085-1086

The most notable innovation and the central idea of the EBM movement was that evidence can be ranked by its quality. For distinguishing between low, good, better, and the best kind of evidence, the EBM movement had to develop a framework fit for such a job. That

is, it had to impose some rules for assessing evidence. In the EBM manifesto, the authors write: "understanding certain rules of evidence is necessary to correctly interpret literature on causation, prognosis, diagnostic tests, and treatment strategy" (Evidence-Based Medicine Working Group, 1992, p. 2421). For example, the quality of evidence refers to the accuracy and stability of causal claims concerning the risk factors for diseases on the one hand, and the accuracy and stability of predictive claims about the interventions to achieve patient-relevant outcomes on the other.

If evidence can be graded, then there is a hierarchy of evidence. This is probably the most well-known contribution of EBM. There are numerous different hierarchies corresponding to different questions we might be interested in. For example, there are hierarchies of evidence concerning questions of treatment, prevention, diagnosis, prognosis, or etiology of a disease. Nevertheless, all hierarchies of evidence assume the same principal ideas about what evidence is the best, what has the lowest quality, and what lies in between.

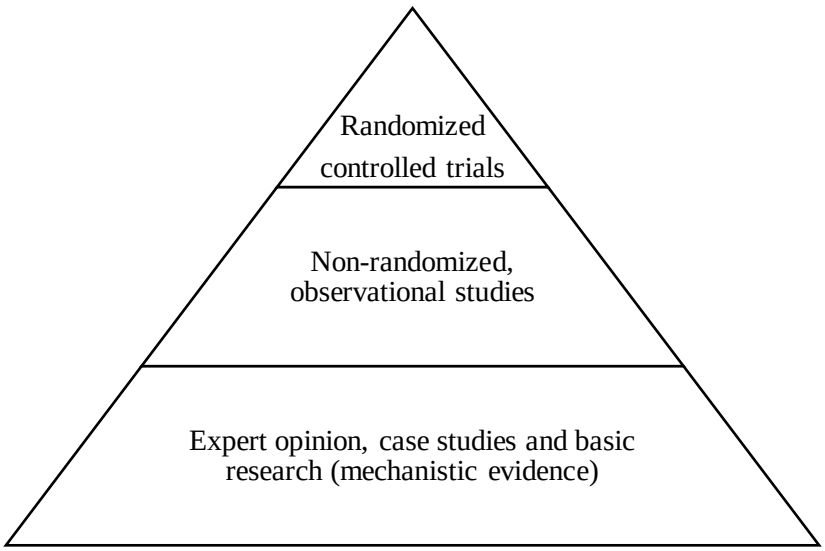


Figure 4. A simple representation of different hierarchies of evidence, as in, for example, Guyatt et al. 2002 or OCEBM Levels of Evidence Working Group 2011.

EBM is empiricist medicine par excellence. It holds in the highest regard the epidemiological studies and the difference-making concept of causation. There are two types of epidemiological studies: observational (descriptive and analytical) and experimental.

Observational epidemiology is concerned either with the identification and prevalence of the patterns, trends, and health conditions of certain populations or with the thorough examination and analysis of the associations (correlations) between exposures to risk factor (e.g., EH) and specific outcome (e.g., heart failure). Descriptive observational epidemiology measures the rates of incidence of a disease in some population, or the distribution of health determinants. For example, it tries to measure the extent to which hypertension is present in different populations, or the average intake of salt in some population. Descriptive epidemiology therefore measures and describes patterns of health and disease and their determinants in some population relative to geographical and time indices.[16]

Observational epidemiology is also concerned with measures of associations or correlations between exposures and outcomes. We can say that observational epidemiology "extracts" causes from associations. That is, it adds a causal import into observed associational relationships. This approach has constituted much of the past and present of epidemiological studies. For the moment, let me present just two of the major types of observational studies – *cohort studies* and *case-control studies*. A cohort study follows a group of people over time and measures the rates of exposure and outcome within that group. A case control study follows a group of individuals – case group – which have been identified as the bearers of the health outcome in question (e.g., some cardiovascular disease (CVD)) and compares the rates of exposure (for example, EH) to that found among a presumably similar enough group – the control group – which lacks the health outcome in question. It is not that uncommon however, that scientists mix these methodologies, and therefore, perform a case control study within the cohort group.

As an example of such observational studies consider the Framingham Heart Study, a starting point in the observational epidemiology of EH. This was a long-term project founded and governed by the National Heart, Lung and Blood Institute and located in Framingham, Massachusetts. Its first original cohort from 1948 included a random sample of 5209 residents of both sexes aged from 30 to 62 from the town's population. In 1971 the study included its second cohort which was comprised of the children of the original cohort and their spouses. In 2002 the cohort was comprised of the third generation. The first major study findings were

[16] So, for example, consider the numbers for hypertension. In 2010, 31.1.% of the global adult population, that is, 1.38 billion people, had hypertension (Mills et al 2016), while in Croatia the numbers were around 45% of the male population in 2018 (WHO 2018).

published already in 1957 and showed an immense increase in the prevalence of coronary heart diseases and stroke among the cohort group. These and other CVDs were then linked to elevated blood pressure or EH (Kannel et al. 1970, 1971, 1972, 1976). The studies also helped to estimate the incidence rates (the rate of individuals who develop a condition within a given time period). In 1971 investigators analyzed the data and demonstrated the risk of coronary heart disease was linked with systolic and diastolic blood pressures: "After 14 years of cardiovascular surveillance of 5,127 men and women in Framingham, Mass., 492 cases of coronary heart disease have been accumulated, thus making it possible to examine *the relation* of each component of the blood pressure to the development of coronary heart disease" (Kannel et al. 1971: 336). By performing multivariate analysis (a set of statistical techniques for assessment of different, independent effects of various exposures on a single outcome) investigators demonstrated a stronger link of systolic blood pressure to coronary heart disease than between diastolic blood pressure and the same outcome. That is, in contrast to elevated diastolic blood pressure, elevated systolic blood pressure is a risk factor for heart attack. But the exact causal mechanism which connects elevated systolic blood pressure to of a heart attack was unknown (and not of interest to the scientists performing the study).

Experimental epidemiology predominately uses randomized controlled trials. Although rarely used to identify etiological causes of a disease (because of the evident ethical constraints), they are used to determine the efficacy of a new drug or a technique of treatment (e.g., surgery). Woodward's interventionist account of causation (2002, 2003) presents an excellent philosophical treatment and theoretical underpinning of randomized controlled trials. Woodward's interventionism asserts that an intervention *I* on the variable *X* acts as "a switch" on all other variables that possibly cause *Y* (Woodward 2003: 98). In a randomized controlled trial, the treatment is provided only through the trial and not from somewhere else. The randomization of the provision of a treatment (e.g., the administration of a drug) is randomly allocated to avoid selection bias. Secondly, in Woodward's account, *I* is not a direct cause of *Y*, meaning the act of administering the drug does not cure disease, rather it is the drug itself that should cause a positive health outcome. An intervention should not change the value of any variable that does not stand directly in the pathway *I*-*X*-*Y*, and it should be statistically independent of any variable that is a cause of *Y* but is not on this direct pathway that goes through *X*. This is ensured by double blinding the participants and randomization. That is, neither the participants in the experiment and control groups nor the experimenters know who was administered with the treatment. Therefore, a change in the value of a variable *X* (for

example, treatment administration 0 or 1) by the intervention (the administration) changes the value of a variable *Y* (health outcome 0 or 1) *if all else is held fixed*. Notice that the randomized controlled trials treat the pathway from *X* to *Y* as a black box.

As can be seen in Figure 4, randomized studies are always at the top of the EBM hierarchies. These can be blinded, double blinded, or not blinded at all. As noted, in the double blinded randomized controlled trials, subjects in both the control and experimental group, together with the researchers, do not know if they received the treatment or a placebo. Therefore, (double blinded) randomized controlled trials are always esteemed most highly; that is, as they are "the golden standard". When EBM-ers say that "[t]he best ("weightiest") evidence for causation comes from rigorous experiments in humans (i.e., RCTs)" (Haynes, Sackett, Guyatt and Tugwell 2005: 358), they think of randomized controlled trials. RCTs are thought to be the least biased type of an epidemiological study. The double blinded type of RCTs secures the experiment from allocation problems, avoid confounding, and distinguishes placebo effects from intervention/treatment effects. However, even better than a single randomized controlled trial are systematic reviews or meta-analyses of numerous randomized controlled trials. Systematic reviews aggregate the results of various RCTs. Since the result of a single RCT is already an aggregated result of numerous outcomes (outcomes of the individuals participating in the study either as part of the experimental or control group), a systematic review is an aggregate result of numerous aggregated results. Rather than providing an aggregated result, meta-analyses combine these results into a singular measure or value.

Although held in the highest regard, randomized controlled trials, however, cannot always be implemented. There are numerous reasons for this, ethical, financial, methodological, etc. For example, we cannot test the effectiveness of the defibrillation of dysrhythmic heart or CPR for stopped hearts by implementing randomized controlled trials. In cases such as these, observational studies are the next best approach to gather evidence. Usually, cohort studies are a level up the ladder than case control studies. The lowest level in all the hierarchies is always occupied by expert opinions, technical note, or the physiological knowledge and research. Case reports or poorly executed case control and cohort studies usually come between the lowest and the middle levels.

According to the EBM "paradigm" or framework, then, the best and/or good evidence in the EBM view always comes from population studies. This, however, assumes that high quality claims about causal relations are always claims about population properties or relations

between populations and not about individual patients. Should the EBM's conception of disease be taken as a population property? Some authors would say yes. For example, Chin-Yee claims that EBM views "diseases as statistical associations at a population level" (Chin Yee 2014: 921). Even if the epidemiological approach retains a biomedical conception of disease (that is, ultimately, health and disease are exclusively biologically explainable), it offers a different way to study their causes – specifically, disease etiology. This has provided a useful way to investigate the risks (or perhaps causes) of CNCDs, since it completely ignores the details of processes inside individuals (the biological level) and includes investigation of what can be conceived as the joint effects of multiple biological mechanisms. In addition, as we have seen, observational epidemiology offers an additional way to manage diseases. By finding factors that correlate with certain diseases, it offers strategies for prevention rather than treatment.

Davidoff et al. argue that critical appraisal of evidence is useful only if it is put into practice, that is, into clinical work. How, then, should EBM work in practice? How is epidemiological evidence and the hierarchy of evidence used in a clinical setting? Consider the example from the original EBM Working Group paper in *JAMA*:

> A junior medical resident working in a teaching hospital admits a 43-year-old previously well man who experienced a witnessed grand mal seizure. He had never had a seizure before and had not had any recent head trauma. He drank alcohol once or twice a week and had not had alcohol on the day of the seizure. Findings on physical examination are normal. The patient is given a loading dose of phenytoin intravenously and the drug is continued orally. A computed tomographic head scan is completely normal, and an electroencephalogram shows only nonspecific findings. The patient is very concerned about his risk of seizure recurrence. How might the resident proceed?
>
> Evidence-Based Medicine Working Group 1992: 2420

According to the authors, the resident can either proceed with "the way of the past" or with "the way of the future". In the way of the past, the resident is instructed by an expert authority (a senior resident) about the usual procedures and guidelines in such situations. Based on these instructions, the resident's biomedical knowledge of seizures, and knowledge of the patient's pathological state, the resident informs the patient about the probability (although not specified by any number) of seizure reoccurrence and advises the patient on how to behave and what to avoid in the future. The way of the future, on the other hand, takes the resident to visit

the library (nowadays, a computer) to conduct their own research of the literature. The resident enters relevant key words (for example, epilepsy, prognosis, and recurrence) and finds a number of relevant papers on the results of population studies. Now enters the hierarchy of evidence. The resident checks for the type of studies used in the research and searches for the highest in the hierarchy. "If the study wasn't randomized, we'd suggest that you stop reading it and go on to the next article in your search […] Only if you can't find any randomized trials should you go back to it" (Straus et al., 2005: 118). If there are no RCTs relevant to the problem at hand, the resident is then advised to check for the next best available evidence in the hierarchy.

EBM presents the most rigid and strongest expression of empiricist medicine up to date and quite arguably is the dominant epistemology of contemporary medicine. All of the major medical journals, such as the *British Medical Journal* and the *Journal of the American Medical Association*, accept its primacy as a medical evidential, methodological, and inferential system. Its methodological commitments and philosophical underpinnings have spread to many other disciplines and sciences such as sociology and economics. Today everything sounds better if it is "evidence-based", from medical treatments and healthcare policies to social and political policies.

1.5. A philosophical analysis of two approaches to disease causation

I have presented the main ideas and views of both the biomedical and the epidemiological approaches to disease causation. In this section, I discuss how exactly we should think about them: are they metaphysical frameworks for disease causation, types of epistemology, causal explanations, causal inferences, or something else?

In his historical analysis of disease causation, Codell Carter argues that causation is "ultimately a theoretical relation, so causal claims can never be justified in the absence of a theory" (Carter 2017: 88). Weed, an epidemiologist, claims similarly to Carter that "any method of causal inference will also have connections with theories of disease (e.g., cancer) causation, the logic and epistemology of causal hypotheses, and the ethics of preventive interventions" (Weed 2000: 798). Similar ideas can be found across the medical literature and the literature on the history of medicine. Authors such as Carter and Weed argue that we cannot develop meaningful or plausible criteria for thinking about causal relations in medicine in the

absence of some theoretical framework of disease causation. There are three desiderata of any theory of disease causation found in the arguments by Carter, Weed and similarly-minded philosophers and historians of medicine: any viable and successful set of causal criteria in medicine will have to: (i) define what kind of things can be causes, (ii) explain how those things figure in some specific theory of causation which should apply to disease causation too, and, finally, (iii) provide methodological and epistemological criteria for the identification of these causes.

There is no doubt that research on disease causation differs in epidemiology and in basic sciences. As already stated, the statistical and mathematical notions, concepts, and heuristics of investigative methodologies in observational and experimental epidemiology differ from the usual laboratory work associated with the investigative methodology of basic sciences. Some philosophers and philosophically inclined scientists argue that the different investigative strategies of causal inquiry in medicine yield two completely different notions of causation and *ipso facto* two different theoretical frameworks of disease causation. Consequently, it is claimed that the epistemological criteria of causality will depend on these frameworks.

At least since Russo and Williamson's paper from 2007, these two notions of cause or causal frameworks have been consistently linked in the philosophy of medicine to a disambiguation of the concept of cause given by Ned Hall. In a nutshell, in his (2004) Hall argues that there are two distinct concepts of cause in philosophy and that most theories of causation in philosophy can be seen as accepting one or the other concept. Hall starts his analysis by laying down five theses about causation that are sometimes overtly and sometimes covertly accepted in philosophical discussion. The first thesis - *transitivity* - asserts that if *a* is a cause of *b* and *b* is a cause of *c*, then *a* is a cause of *c*. The *locality* thesis asserts that causes and effects should be spatiotemporally contiguous, that is, they are connected by series of spatiotemporally connected causal intermediaries. The *intrinsicness* thesis claim that every causal relation is determined by its structure which is non-causal in character (together with the laws of nature). The *dependence* thesis asserts that counterfactual dependence between distinct events is sufficient for causation. The fifth thesis, *omission*, claims that absences and failures of event occurrences can both be causes and effects.

Hall attributes the first three thesis to the productive concept of causation, whereas the second two are linked to the dependence concept. Hall then proceeds by considering different

scenarios of purported causal situations where these five theses cannot all be held at the same time. Hall argues that the dependence accounts of causation have trouble incorporating the first three thesis, if pressed with stubborn and persistent counterexamples of overdetermination. The acceptance of the first three thesis cannot be reconciled with the acceptance of the fourth and the fifth thesis, as the counterexamples that involve the double-prevention scenarios vividly present. These two distinct concepts capture our two confronting intuitions about causation and what a cause is supposed to be.

Although Hall speaks of dependence, some authors have claimed that the terms "causal relevance" (Glennan 2009, 2017) and "difference-making" (Russo and Williamson 2007, Illari 2011, Clarke et al 2013) mean the same thing. Hall takes this concept to be a counterfactual dependence relation between two events in the style of Lewis' back-tracking counterfactuals. But since then, philosophers have argued that the dependence concept or difference-making concept accommodates interventionist (Woodward 2002, 2003) and probabilistic (Reichenbach 1956, Eells 1991) accounts of causation; that is, any theory of causation that takes a cause to be a difference-maker for the occurrence of their effects (and where causation is then analyzed by providing some truth conditions for the difference-making relation). For present purposes, the most important features of these theories are that causes are states, events or variables, and that causes are difference-makers in the sense that they make a difference to either (i) whether the effect will or will not occur, (ii) the probability that the effect will occur conditional on the cause occurring rather than not, or (iii) the magnitude, scope or frequency of the effect occurring. In another words, difference-making causality is a measure of dependence between values of variables (whatever the variables stand for), when certain theoretical conditions are satisfied (for example, temporality or independence of variables in interventionist account). Many philosophers have taken that the difference-making concept of causation corresponds to causal rationale in epidemiological studies (both observational and experimental). The difference-making concept then designates EH as a cause of heart failure because there is a counterfactual dependence relation between the two or because hypertension severely raises the probability of heart failure (based on population-level observations).

Hall takes the productive concept of causation as far more difficult to capture than dependence, but intuitively, he argues, we take it as a matter of causes producing, bringing about, or generating their effects. There is a tendency in philosophy to take the notions of production or bringing about as non-reductive notions. For example, Anscombe (1973)

famously argued that the notion of cause is ambiguous and vague. It only acquires some meaning when it is substituted for some more specific term, such as bonding, pushing or pulling. Hall, however, does not claim that production is (necessarily) a primitive, non-analyzable notion. He offers an attempt to lay down a reductive (in the sense of reduction to non-causal terms) analysis of production. Production, then, is a matter of having the right kind of internal structure which he identifies as a union of minimally sufficient sets for *e* in every time between *t* and *t'* – (the time of *e*'s occurrence).[17] This would mean that *X* (having essential hypertension) produces *Y* (having a cardiovascular disease) if it can be shown that these events are linked by a series of steps, each defined by its minimally sufficient structure of conditions, relations and objects. In other words, Hall's notion of production corresponds to having a unique structure of events or objects which are organized so as to eventually lead to or bring about the effect. *X* produces *Y* when there is a spatial and temporal organization of entities, their interactions and activities connecting *X* and *Y*. Similarly to the case with difference-making and the epidemiological approach, philosophers have argued that the production concept corresponds to the biological or mechanistic approach to disease causation. *X* is not a cause of *Y* only because, statistically or counterfactually, it is a kind of difference-maker to the occurrence or magnitude of *Y*. Rather, *X* is a cause of *Y* because there is a certain biological mechanism or mechanisms of the right sort connecting the two events or states and which can be traced or split into stages where each stage is at least sufficient to bring about the next stage because of the entities involved, their properties and relations.

Perhaps a better fit for the topic of this dissertation is Woodward's analysis of these two views. That is, although claiming the same thing, his presentation of those views is closer terminologically and in its focus on particular details to the discussion in the philosophy of medicine:

> While [difference-making] accounts assign a central role to contingency information as a source of evidence for causal claims, [mechanistic] accounts commonly assign a central evidential role to spatio-temporal or geometrical relationships or to facts about the presence or absence of the appropriate sorts of mechanical properties (rigidity, weight etc.). In particular, cases in which (it appears) one can just "read off" which causal relationships are present from geometrical or mechanical properties, without any

[17] Observe the similarities between Hall's definition of a production and Mackie's and Rothman's definitions of sufficient conditions.

> apparent need to rely on contingency information, play an important role in [mechanistic] thinking about causation.
>
> Woodward 2011: 413

Whether or not Hall's arguments imply that not only are there two different concepts but also two different *kinds* of causal relation is a matter of debate and Hall remains silent on this question.[18] Nonetheless, in the next two sections I present how this philosophical discussion on two concepts of cause is reflected in medicine's two approaches to disease causation. I argue that the discussion on difference making (relevance and dependence) and the productive/mechanistic causal frameworks in medicine should be understood as different kinds of causal explanations rather than different metaphysical concepts of causation. The difference lies in the content of explanation, the methodology by which those explanations are arrived at, and the evidence which is gathered for their support. I will start with the argument that these two are different forms of causal explanations rather than different metaphysical theories of causation, and then discuss the evidence provided to establish causation in medicine.

1.5.1. Explanation

Recall Henle's quote from the section 1.1. In addition to the claim that there is a high correlation between falling bodies and unsupported bodies, physics, as Henle argues, can say *why* bodies fall without being supported. Along these lines, the standard correlation approach of modern epidemiology has been criticized for lacking any explanatory relevance. It is often argued that the statistical inquires of epidemiological observational and experimental studies only reveal that, for example, hypertension is highly correlated with the lack of exercise, cholesterol-rich or salty food or tobacco and alcohol consumption, but this does not say anything about how or why smoking or salty food increases the risk of developing hypertension. We would like to know why and how changes in one variable produce changes in another, not just that they do so. In other words, the difference-making or black box approach only "leads to the identification of a list of risk factors", and therefore, it is not "an explanatory

[18] See, for example, Russo and Williamson (2007) and Strevens (2011) for the arguments that causation just is one relation but there are different kinds of evidence for causation, Glennan (2009), (2017) for the argument that there are, in fact, two kinds of relation, and Cartwright (2004) that there are multiple different concepts of cause, corresponding to different causal questions and methodologies.

theory for how disease arises" (Hafeman and Schwartz 2009: 838). In addition, the values of variables and variables that we measure either by observation or by interventions must represent something from "the real world". We have to have some idea what the things that we measure are.

There are two standard criticisms of the difference-making approach (both as a theory of causation and as an explanatory approach in medicine) which are supported by the first two desiderata of disease causation presented at the beginning of the previous section. Although useful in detecting numerous determinants of health outcomes in some population, the difference-making approach cannot give answers to two related questions: (i) what sort of things can be causes (since we can measure whatever we want, for example the relation between exercise and school achievement), and (ii) how we arrive at causal claims from measuring correlation (but, even if we can do this, how exactly the exposure causes the outcome is still a question without an answer).

The difference-making or black box stance and the statistical evidence which it relies on have been criticized for putting too much emphasis on the pragmatic aspects of finding causes or correlations. This does not come only from academic philosophical discussions which are not recognized by epidemiologists themselves. These sentiments have been present in discussions and articles in epidemiology and they concern many epidemiologists (for example, see Vandenbroucke (1988), Skrabanek (1994), Hafeman and Schwartz (2009), to name a few). Causal explanations, then, are perhaps completely and intentionally lacking in epidemiology. This led some prominent epidemiologists to be wary when it comes to assessing causal claims. Others went a step further and denied the validity of any causal talk whatsoever. For these causal deniers, epidemiologists can only talk of associations and associations cannot give us causal relations (whatever they may be). For example, observe the passage by Lipton and Odegaard in their (2005).

> Our point is that although it is important to be able to use epidemiological research to predict and intervene at the public health level, to tell the best story possible about the research findings at hand, one doesn't have to say that *X* causes *Y* to achieve such an outcome. In fact, one cannot definitively claim such a relationship.
>
> Lipton and Odegaard 2005: 7

Because of the issues with the notion of cause in epidemiology and its discovery of causal mechanisms, a great deal of traditional epidemiology has tended to completely abandon casual terminology in favor of a "broad and nonspecific category of *determinants*" (Susser 1991: 637). Rather than getting into the problem of stipulating what a cause is and what its properties are, many epidemiologists have concentrated on getting the most out of its methodology without making explicit causal claims. "Many public health researchers were taught that it is best to avoid discussion of causation in interpreting findings from observational studies; any reference to causation was thought to overreach the evidence" (Glymour and Spiegelman 2017: 81). This avoidance of causal talk certainly has its positivist and logical positivist roots (Kincaid 2011). From these scientific and philosophical standpoints observational and experimental epidemiology often accepted a view on causation as too metaphysical a notion which, if possible, should be avoided in scientific theories and models.

Many epidemiologists, however, are sincerely worried that statistical methods and concepts cannot be the end point in the investigation of diseases. To explain a phenomenon or a pattern citing strong or resilient correlations between variables is not enough. Even leading epidemiologists have recognized this insufficiency and expressed their concerns. For example, Mervyn Susser argues that association is only one "sine qua non" feature of a cause but not the only one and certainly, far from sufficient to establish causation. Susser writes: "If no grounds for an association can be shown to exist, causality has been rejected, and we proceed no further" (Susser 1991: 638). What are the grounds for association? Susser mentions a few (for example, time order and directionality) but still considers them as insufficient to establish causation by themselves. The association, however, can be further bolstered by other statistical measures and concepts, so some epidemiologists and statisticians would object here and claim that the causal interpretation of an association can be reached by implementing carefully designed studies. However, the problem is that further associational or correlational measurement do not "change the fact that the thing being directly evidenced is an association" (Fiorentino and Dammann 2015: 3). The motto "correlation is not causation", or in Cartwright's variation "No causes in, no causes out", is a shared feature of statistical sciences, epidemiology included (Cartwright 1989). Causal models are always associational models, but associational models are not always causal models. So, what does it mean to provide a causal explanation of statistical measures? What does it mean to provide a causal interpretation of measures such as risk ratio or relative risk?

If causal investigation and causal explanation mean providing answers to questions about *how* associations or correlations are instantiated, as mechanistic philosophers and mechanistically minded scientists would argue, difference making accounts seems badly suited for such a task. As Clarke and Russo (2017) claim, difference-making says *that* the risk factor correlates with the outcome, but we also want to know *how* it makes a difference. Causation, or a causal reading of association, therefore, is not found in the relations being measured but rather it is derived from some theoretical background – specifically, a background in biological theory, as the biomedical advocates to disease causation would argue.

Consider the following question. Can the difference-making approach of epidemiology offer an answer to the question of what connects variables we choose to intervene upon and measure? No. The interventionist and probabilistic accounts of causation, which figure prominently in both observational and experimental epidemiological studies, do not offer an answer to this question. This is recognized by the proponents of the interventionist account of causation in philosophy: "Even if the [interventionist account] does not identify the truth-maker for causal claims, it is nonetheless an illuminating analysis of the causal truths themselves, and it is crucial for the project of deciding which putative metaphysical explanations (that is, which truth-makers) are adequate and which are not" (Craver 2007a: 106). The role of biological models and the knowledge of biological mechanisms is not only diminished in epidemiological studies (except in molecular epidemiology) but ignorance about the biological mechanisms is sometimes even explicitly endorsed among black box proponents.[19]

One possible strategy is to take a purely pragmatist position and claim that the search for underlying biological causes can delay the acceptance of discoveries arrived at by the difference-making approach. Therefore, we are far better off embracing the difference-making approach of strict associations and without causal claims when talking about medical treatments or about the factors that influence the onset of diseases. For example, Semmelweis had compelling evidence (both observational and experimental) in favor of the causal hypothesis of association between "cadaveric material" and puerperal fever yet his contemporaries were reluctant to accept it because, among other things, he did not identify the biological mechanism and the biological theory of the day – the miasma theory – did not allow

[19] For example, the title of Richard Peto's paper from 1984 is "The need for ignorance in cancer research".

for it.[20] The neglect of biological theories in the experimental and observational methods of the difference-making approach was praised even in philosophy (Ashscroft 2004). Biological considerations on the plausibility of a causal association can be in direct conflict with the observational or experimental hypotheses. If Semmelweis had made a randomized controlled trial to test the effectiveness of hand washing, it would have been shown that such an intervention considerably lowers the incidence of puerperal fever. Even if everyone else held the miasma theory, his critics would have had a hard time to oppose the findings of the experimental method. Biological theories and models are fallible and there is no guarantee that what holds today will hold tomorrow. Recall how the fallibility of biological reasoning was one of the reasons for EBM to arise in the first place. The difference-making or statistical (epidemiological) approach is just more reliable (if the studies are done correctly) than inferring predictions from laboratory research, irrespective of whether or not we can claim causation.

Although I am sympathetic to these considerations, I will not discuss them further. I will, however, provide an argument in favor of the utility of the difference-making approach but only because it is, in a way, a provisional causal explanation.[21]

Knowing how something makes a difference will not always be the information we seek. Sometimes the answer to a question requires includes knowing that it makes a difference. That is, "finding how" and "finding that" result from asking different questions. The epidemiologist Geoffrey Rose defines these different questions as questions concerning case-studies and questions concerning incidence rates: "'Why do some individuals have hypertension?' is a quite different question from 'Why do some populations have much hypertension, whilst in others it is rare?'" (Rose 2001: 428). Although we do not know a lot about the mechanisms connecting obesity and hypertension, the connection itself is fairly well understood. We can use it to provide a token explanation – an explanation of why an obese individual *X* developed hypertension. Similarly, consider the example discussed by De Vreese et al. in (2010) about skin cancer incidence in two groups of Belgian tourists, one that spends their holidays in the Mediterranean and one that spends them in Belgium. Why is skin cancer

[20] The most well-known philosophical discussions on causal reasoning in the Semmelweis case are in Hempel (1966) and Lipton (2004). Semmelweis reasoning has been taken as an example of a good causal reasoning. However, for a different, critical interpretation of Semmelweis's causal reasoning see Tulodziecki (2013).

[21] Although not directly discussing the issue of epidemiological causal explanations, Glennan's "bare causal explanations" from his (2017) can be thought of as expressing the same idea.

more prevalent in the first group? The difference in holiday location explains the prevalence. But this is a population property of a subpopulation of Belgians that spend their holidays in the Mediterranean. On the other hand, there is fairly well understood causal process connecting exposure to UV radiation and the pathogenesis of skin cancer. If a particular Belgian *X*, who spent her holidays in the Mediterranean and never used sunscreen, developed skin cancer, we could give a good causal (mechanistic) explanation for it. These different questions about populations or about specific cases require different methods and different concepts. Following this, De Vreese et al. (2010) claim that medicine offers different kinds of explanation: we can give explanations involving features of populations, of individual(s) comprising those populations, or the components of those individuals. Only the last of these kinds of explanation require the mechanistic approach to disease causation.

I agree with De Vreese et al., and defend the view that both approaches are explanatory, but they offer different kinds of explanations which provide answers to different questions. That is, they represent different strategies of inquiry to answer different aspects of the same medical phenomena; namely, (i) questions concerning the population level and intra-individual level, and (ii) questions concerning etiology vs. pathogenesis of disease. To answer these different questions, different methodologies, evidence, and notions of causal explanation are needed. I will discuss (i) and (ii) here and methodology and evidence in the next section.

First, observe how Rothman defines the distinction between risk factors for individual and for populations:

> For an individual, risk for disease properly defined takes on only two values: zero and unity. The application of some intermediate value for risk to an individual is only a means of estimating the individual's risk by the mean risk of many other presumably similar individuals. The actual risk for an individual is a matter of whether or not a sufficient cause has been or will be formed, whereas the mean risk for a group indicates the proportion of individuals for whom sufficient causes are formed. An individual's risk can be viewed as a probability statement about the likelihood of a sufficient cause for disease existing within the appropriate time frame.
>
> Rothman 1976: 589

When person *X* has a heart attack, it means one of the sufficient component sets has obtained. We can then study what the factors were in this sufficient component set. But heart

attacks in a population of which *X* is a member might more highly correlated with different factors or sufficient component sets. That is, what is significant in an individual case is not necessarily what is significant for the population.

Consider the population *P* and one its member, the individual *X*. *X* just had a heart attack, and we are curious why *X* had a heart attack. *X* is living a rather unhealthy lifestyle. *X* is an avid smoker and does not exercise. *X*'s smoking and lack of exercise seem like a major risk factor for *X* to have a heart attack. Eventually, *X* has a heart attack. On average, members of population *P* live similar lifestyles to that of *X*. Consumptions of tobacco and alcohol are high in that population and we have measured a high correlation between heart attacks and consumption of tobacco and alcohol in that population. So, what is the cause of *X*'s heart attack? - Lack of exercise and excessive tobacco consumption. We cite the characteristics and behaviors of *X* and not his biological parts.

Weed takes it that the black box should be taken as a metaphor for an individual so that the black box stance is a "methodologic approach that ignores biology and thus treats all levels of the structure below that of the individual as one large opaque box not to be opened" (Weed 1998: 13). This is certainly informative, but it is not an explanation in terms of biological causes (what does, metaphysically speaking, lack of exercise even mean?). A physician then might say that the cause of *X*'s heart attack is the obstruction of the blood flow to the heart due to blood clot which, on the other hand, happened because of smoking induced atherosclerosis. The physician (or pathologist maybe) seeks to understand how the *X* heart attack came to be in terms of individual *X*'s parts. This requires "parsing an individual in terms of his or her biologic make-up rather than externally observable characteristics and behaviors" (De Vreese et al. 2010: 374). Similarly, Fiorentino and Dammann proposed in their (2015) that difference-making approach measures the correlations between variables and then, by using different statistical tools, proposes causal hypotheses to account for these patterns. The mechanistic stance, on the other hand, offers a biological explanation of these hypotheses. It does so by referring to its intra-individual manifestations in terms of biological causes. In another words, the difference making or black box approach stops at the level of the individual, while the mechanistic approach, goes into the individual and looks for the processes occurring on the intra-individual level.

Consider now a different population – *P**. On average, the population *P** is living a healthy lifestyle with lots of exercise, healthy dietary habits with low alcohol, and no tobacco

consumption. Heart attacks in that population are rare. The major risk factor for the incidence of heart attacks in that population is not excessive tobacco use. Let us say that the major risk factor for a heart attack in that population is inherited, i.e., genetic. If we want to know why population *P* has more heart attacks than population *P**, biological information provided by the physician above will not be informative. But comparisons in relevant characteristics between groups will give us an answer. The incidence of heart attacks in population *P* is higher than in population *P** because tobacco and alcohol consumption in population *P* is much higher than in *P**. Tobacco and alcohol consumption are difference-makers.

To conclude, both approaches are causally explanatory but only if seen as answering different questions: questions about causal relations on the population vs. intra-individual level, and questions about etiology and pathogenesis of diseases. Consequently, these two approaches are based on different evidential support. However, the approaches do not imply different metaphysics of causation. The causal theory by which we describe relations between entities composing mechanisms at the intra-individual level does not need to be different from the one in the difference-making approach. Many mechanistic philosophers take causal relations between mechanism's parts along the line of the interventionist account of causation (e.g., Glennan 2002, Tabery 2004, or Craver 2007a). I also defend a similar claim in the next chapter where I argue that mechanistic explanations of some biological phenomena need both concepts of causation.

1.5.2. Evidence

In their (2007) Russo and Williamson argue that causal claims in medicine are accepted or warranted only when we have evidence of both the difference-making relations or statistical evidence and the evidence of a mechanism connecting exposure and outcome.[22] That is, in addition to knowing that in some population heart attacks are highly correlated with tobacco and alcohol consumption we have to know how such behavior causes heart attacks on an intra-individual level. Having only one type of evidence is not enough to warrant a causal claim and strictly speaking, there should not be two different types of causal explanations. This has

[22] In the paper from 2007, Russo and Williamson talk about probabilistic evidence, while later, in (2010), they accept the notion of difference-making to include probabilistic, statistical, or epidemiological methods and evidence.

become known as the "Russo-Williamson thesis" (RWT). If RWT is a plausible claim then it is a counterargument to my claims from the previous section: namely, to answer *any* question concerning disease causation one needs both types of evidence and *ipso facto* implements both strategies.

What, then, are the kinds of evidence for difference-making relations and the evidence for mechanisms? Illari (2011) and Bluhm and Borgerson (2011) claim that epidemiological observational studies (cohort and case-control studies) and experimental studies (RCTs and laboratory experiments) are not themselves *the* evidence but rather evidence-gathering methods. They are different types of studies and experiments, that is, methods used to gather data. Therefore, we could claim that both RCTs and observational methods gather the same evidence – namely, relations of associations among groups. On the other hand, different laboratory experiments are supposed to reveal biological and chemical mechanisms. The data, it may also be assumed, can be quantitative or qualitative.

However, both difference-making and mechanistic evidence can be gathered by observational and experimental studies. No matter the studies we perform, we will be looking for the features of difference-making relations and dependencies between variables revealed by observational and experimental epidemiology and mechanistic structures and processes found through the work of medicine's basic sciences. In that regard, the evidence that these methods reveal are a mark, a sign, or an indicator of causation but not the same thing as the causal relation (for a similar view on evidence for causation see, for example, Reiss 2015).

To illustrate what Illari has in mind when she claims that features of evidence are the distinctive mark of difference-making relations and mechanistic relations, I will use Bradford Hill's influential criteria for the assessment of causation from association (Hill 1965). In a famous and influential paper from 1965, Hill proposes nine criteria or features of causation that a researcher should address when deciding whether the association is due to causation. Hill does not argue that any specific criterion is necessary to warrant a causal claim. In fact, he does not think that satisfaction of all nine criteria guarantees a causal claim. Assessment of any causal claim is conditional upon the peculiarities of a specific study or case. Nevertheless, those criteria are supposed to be good indicators that an association between variables could indeed be causal. More importantly, as Russo and Williamson (2007) observe, the criteria nicely illustrate difference-making and mechanistic evidence. Let us examine them in turn.

1) Strength – If the strength of an association between two variables is strong it will be more likely that it is due to a causal relation rather than to some additional variables controlling their values. For example, the rate of cardiovascular diseases is much higher among people suffering from essential hypertension than among those that do not.
2) Consistency – Claiming a causal interpretation of an association seems more likely if different methods of scientific inquiry (cohort studies, case-control studies, randomized controlled trials etc.) give us similar results across different populations (males and females, or populations in different countries)
3) Specificity – One should look for a single cause of an effect (one cause, one effect). Yet, as we have seen in section 1.3., specificity seems to be too demanding a criterion, especially in the research on causation of CNCDs. It might be considered that lung cancer prevalence among smokers is high enough that we should conditionally consider smoking as *the* cause of lung cancer (Russo and Williamson 2007). However, cardiovascular diseases still seem to be too multifactorial to have the criterion of specificity carry the same amount of evidential force as other criteria on the list.
4) Temporality – Simply, causes ought to precede their effects in time. The onset of a disease is temporally posterior to its causes. Although a reasonable metaphysical condition, the temporality criterion is not so immediately clear in the cases of CNCDs. As Hill comments: "Does a particular diet lead to disease or do the early stages of the disease lead to those peculiar dietetic habits?" (Hill 1965:9). In the early days of the investigation into the relation between smoking and lung cancer, scientists had to rule out the hypothesis that people with lung cancer start to smoke or start to smoke more often when they have found out about their condition. In the case of hypertension, it is still not settled whether the endothelial dysfunction happens because of hypertension and is temporally posterior to it, or whether it is a cause of hypertension. Therefore, the presence of endothelial dysfunction does not necessarily imply that we should expect hypertension to follow.
5) Biological gradient – Although it appears originally in Hill's paper as a "biological gradient" the fifth criterion is maybe better termed as a "dose-response relationship or curve". It states that cause-effect relations are more likely to represent a linear (or perhaps even more likely exponential) relation between the dependent and independent variable: for example, the prevalence of hypertension rises with the daily intake of cholesterol rich food.

6) Plausibility – In other words, this criterion implies that cause-effect relations should be biologically plausible. The suspected causal association would gain considerably more weight if we had an underlying biological mechanism identified or at least if we suspected some possible mechanisms. For example, the hypothesis that hypertension is a risk factor for heart attack is more likely to be true when we know that suffering from chronic hypertension results in narrowing the coronary arteries by the accumulation of fat and cholesterol. This eventually leads to the formation of blood clots which prevent the supply of oxygen and other nutrients to heart. The result of these processes is myocardial infarction – heart attack. However, Hill notices that what is biologically plausible will be dependent on the biological knowledge of the day and this can often contradict associational findings. There is no better example of this than the case of the reluctance of the medical mainstream to accept Semmelweis' hypothesis.
7) Coherence – Similarly to the previous criterion, the coherence criterion tells us that the assumption of a causal association will be more likely if it "makes sense" or it is coherent with our body of knowledge. This does not necessarily imply biological knowledge. Closely related associations can make an association between *X* and *Y* more coherent. The hypothesis about hypertension being a risk factor for heart failure is more likely if there is an observed high correlation between hypertension and heart attack.
8) Experiment – Evidence gained from experiments greatly contribute to the plausibility of a causal interpretation of an association. The experimental evidence can be gained from randomized controlled trials as in the cases with experiments with drugs used to treat hypertension but also in laboratory rats or other animal models where a mechanism or an association can be reproduced. The shared feature of both types of experiment is the intervention by a researcher. As I will show, laboratory experiments performed to reveal underlying mechanisms often follow the same causal epistemological rationale as randomized controlled trials – namely, Woodward's interventionist account.
9) Analogy – Analogical thinking by using animal models or analogical thinking by pointing to similar causal mechanisms and pathways makes us more confident that the assumption under consideration holds.

The evidence for difference-making relations is that changes in the purported effect variable vary in accordance with changes in the cause variable, or that changes in the outcome variables move accordingly with changes in the exposure variables – e.g., smoking and lung cancer. In this sense, strength, consistency, dose-response relationship (biological gradient),

and experimental evidence gathered from randomized controlled trials are evidence that the exposure (e.g., to a pathogen, working or living environment, nutrients) is a difference-maker to the outcome of interest – remove the exposure and the incidence of the outcome ought to drop. Temporality, plausibility, coherence, experimental evidence gathered through laboratory work and analogy are evidence of an underlying mechanism or a process linking exposure and outcome.[23]

Why do Russo and Williamson claim that these two types of evidence are inconclusive on their own? Let us consider the evidence of difference-making relations first. In epidemiology, the existence of an association has three potential explanations: an association between two variables arises due to (i) a causal relation, (ii) a confound, bias, or some other error in the design of the study, or it is a matter of (iii) a chance, an accident. Before accepting the causal interpretation of an association, scientists need to rule out (ii) and (iii): "The rationale supposes that differences in probability need a causal explanation, and, if all explanations relying on confounders are eliminated, then T causes O is the only explanation left" (Cartwright 2007: 15). What makes a biased study? Observational studies are commonly more prone to these errors, but they can be found in the (double blinded) randomized controlled trials as well. For example, confounding (or common causes in philosophical jargon) implies that there is a hidden factor which controls for both of the variables, and which has yet not been detected. Fisher (1958) famously doubted the assumption that smoking causes lung cancer based on the observational evidence because researchers had not eliminated a possible confounding factor – a gene responsible both for lung cancer and smoking addiction. A study follows some proportion of people – a sample of a population. But how are they selected? It is up to researchers how the groups will be selected. Differences in the selection of groups in both observational and experimental studies can lead to numerous selection biases or allocation biases. Researchers can choose people in an experimental group who are in a more serious condition, or who they think will respond more positively to treatment (they can also fail to recognize that the group(s) selected for the study will, because of some feature, respond better or worse to treatment than the general population from which they were selected). Double blinded randomized controlled trials are not immune to the same problems that burden non-randomized controlled trials or observational studies. Confounding factors are never ruled out

[23] As mentioned previously, it could be that the specificity criterion presents an inadequate criterion considering the etiology or pathophysiology of CNCDs so the specificity criterion is better taken as optional or a relic of a once influential view in medicine that is not so well adapted for present purposes.

unless we have sufficiently large groups in the study which, of course, we never have (what would that be?). Another common problem is that scientists sometimes decide to measure surrogate outcomes instead of the desired ones – for example, instead of measuring how much mortality is reduced, scientists decide to measure some other variable they find convenient.[24]

How can the knowledge of mechanisms provide the epistemic and metaphysical grounds to infer causal claims from associational relations? Similar to Weed and Hursting's discussion in their (1998), I identify three ways in which the evidence of mechanisms can provide the rationale for the conclusion that the correlation is causal, and therefore, provide a causal explanation of the correlation. The first way, and the strongest form of the mechanistic stance, requires knowledge of the actual, real mechanism that underwrites the causal connection between A and B. This form of the mechanistic stance would claim that it is not enough that there is a plausible or possible mechanism or that there are several plausible and possible mechanisms. We are in position to offer an explanation and give predictions only when we know the actual mechanism responsible for the phenomenon. This kind of argument is usually propounded by the new mechanistic philosophers. For example, Machamer, Darden and Craver (2000), Woodward (2002) and Craver and Darden (2013) repeatedly make such claims: "However, if one's goal is to control, explain, and/or predict how a mechanism will behave under the widest variety of conditions, one requires more than a mere how-possibly schema. For those purposes it is often crucial to know how the mechanism *really works*" (Craver and Darden 2013: 34, emphasis added). A causal claim or hypothesis, in that case, will be justified or warranted when the mechanism that produces or is responsible for the phenomenon is identified - meaning that all components (and their properties, activities, and interactions) and their causal, spatial, and temporal organization are identified and understood. This amounts to the claim that to make a causal hypothesis in biomedicine it is not enough to have sufficient evidence that something works, rather we must know how it works and that indeed it does work in that manner.

However, the demand for a singular mechanism underlying an association cannot be universally applicable. This will be a matter of empirical investigation rather than a theoretical presumption. It is not that uncommon that *X* and *Y* are causally connected by several different mechanisms. For example, in his (1999), Thagard discusses how the medical community came

[24] See Howson and Urbach (1993), Worral (2002), and Howick (2011) for detailed discussions on these issues.

to accept that *H. pylori* causes ulcers. To accept this causal claim scientists gathered both the statistical, difference-making evidence of the presence of bacteria in patients with gastric ulcers and the mechanistic evidence that explained how the bacteria cause ulcers and how it is possible for it to survive in the inhospitable environment of the stomach. But this is not the only mechanism by which gastric ulcers can arise. Nearly 20% of all ulcers are caused by non-steroidal anti-inflammatory drugs such as aspirin, ibuprofen, or naproxen and 1% of ulcers are actually the result of the Zollinger-Ellison syndrome where the body produces excessive amounts of acid. In addition, a lot of people carry the *H. pylori* bacterium in their stomachs, without getting gastric ulcers. Similarly, essential hypertension, as it was mentioned in previous sections, can be caused by a number of mechanisms. What does an actual mechanism mean in such a case? There are numerous mechanisms which lead to developing chronic high blood pressure and it is quite possible that several of them can be simultaneously at work.

Hence, a less demanding reading of the mechanistic stance is perhaps more suitable. Such a reading would require only that, presumably, there is some biological or social mechanism which can form a link between the two variables that have been observed to correlate. This is shown in the Figure 5 as the "Plausible mechanism M1" or "Plausible mechanism M2". We might know that there are several mechanisms which can plausibly produce the effect but there might be no conclusive evidence as to which mechanism produces the effect in general or produced the effect in a specific instance.

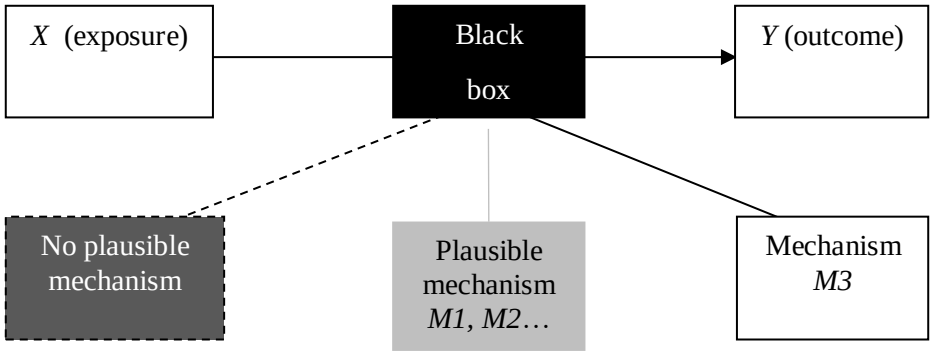


Figure 5. A simple representation of the three aspects of the mechanistic interpretation of association between exposure and outcome.

An even less demanding reading of evidence of mechanisms can sometimes be applied to corroborate or discard a causal hypothesis. This would include negative mechanistic inference where we either temporarily withhold judgment about whether the association is causal since no mechanism has been found, or completely disregard the causal assumption simply because no plausible mechanism is conceivable (Jerkert 2015). The first option is what Broadbent has in mind when he claims that the mechanism for the operation of some risk factor must be identified eventually: "Perhaps not immediately; but if in the fullness of time no mechanism is identified, the credibility of the hypothesis will suffer" (Broadbent 2011: 49). A nice illustration of the second option, that is, the "no plausible mechanism" solution, is Thagard's example of homeopathic medicine. Homeopathy's assertion that mild doses of drugs diluted in water have therapeutic effects conflicts with everything we know about chemistry, biology, and physics. Simply, there are no plausible mechanisms which could make the homeopathic preparations effective in treating or preventing diseases.

According to Russo and Williamson (and Illari (2011) too), evidence of mechanisms is also insufficient on its own to make a conclusive causal claim in medicine. Consider again the example of multiple mechanisms in the development of hypertension. Even if we find a mechanism which potentially can bring about the outcome this does not imply that it indeed brings about the specific outcome. There could always be another mechanism which produces the outcome – in other words, mechanistic explanations are prone to the masking problem. Also, there could be several different mechanisms for the development of hypertension at work at the same time. How much of the effect then is due to a particular mechanism? Furthermore, the very same mechanism which produces an outcome or increases the value of an outcome variable can in different setting prevent or decrease its value. For example, Steel (2008) mentions the simple example of exercise and weight loss. Exercising leads to the burning of calories and therefore, weight loss, but at the same time it raises one's appetite which again leads to a higher intake of calories. By observing the behavior of a mechanism in one setting we cannot have a ready-made answer to how it will behave in another setting. Although oversimplified, similar examples are regularly present in various biological mechanisms.

Although a simple argument, RWT ignited a lively and interesting debate in the philosophy of medicine (but science in general too) with some of the authors agreeing to some extent (e.g., Illari 2011, Gillies 2011, Claveau 2012, Clarke 2013) while others have argued

that it is plainly wrong (e.g., Howick 2011a, Reiss 2012). So, is RWT a plausible thesis? If we take RWT in its strong reading, it is evidently wrong both as a descriptive and normative thesis.

Descriptively, RWT just does not reflect scientific practice in medicine. It does not reflect that there are different questions concerning etiology and pathogenesis of diseases which require different explanations and evidence. Concerning treatments, on the other hand, there are numerous causal claims in medicine taken as conclusive even though they have never been tested by RCTs (e.g., defibrillation as the procedure for resetting a dysrhythmic heart) or, where the exact mechanism has not been found (the use of lithium for the treatment of bipolar disorder, or the use of aspirin until the 1970s).

As a normative thesis, however, RWT fails for the reasons I have explained in the previous section. First, it does not recognize etiological vs. pathogenic aspects of disease causation. Does the incidence of skin cancer in the Mediterranean vacationers from Belgium versus the ones that do not go there matter for the scientists trying to come up with an explanation of the pathogenic processes occurring in skin cancer? Of course, one could argue that we would not know that unprotected exposure to strong UV radiation causes skin cancer if we did not have the epidemiological evidence in the first place. The same goes for numerous other etiological relations between the exposure and the outcome. But, again, these are not the same questions. In addition, considering its practicality, taking RWT as a normative thesis would have delayed numerous treatments and health policies for decades. Considering the amount of observational evidence in favor of the claim that smoking causes lung cancer that we have had for decades, it was not that long ago that we have come to understand which carcinogens are present in cigarette smoke and how they cause lung cancer. There are numerous examples from medical science where the mechanism underlying a causal association has been unknown for decades, yet we have exploited the association to satisfy some health-related ends. Uses of aspirin or lithium in medicine are good examples where the mechanisms of action were completely unknown for some time, but their beneficial effects were well-established and more importantly exploited for positive patient-relevant outcomes. Eventually, we have come to understand the mechanism by which aspirin has its analgesic effects and there are several proposed mechanisms of action of lithium in the treatment of bipolar disorder. But the need to find the mechanisms grounding causal associations before such associations can be used for medicinal purposes would is too demanding a condition when we know that the stakes of medicine are the highest. Aspirin was doing a good job of relieving pain before we had any

idea how exactly it does that. We did not observe any counter-indications and it would then seem irresponsible not to use it to treat headaches. You do not need to know anything about the internal combustion mechanism of car engines to know that turning a key, ceteris paribus, starts the car engine. We knew that aspirin on a population level relieves people of pain and have exploited that association. But we could not give an explanation why it did not work in some individual cases. On the other hand, some mechanisms are so plain and simple that so scrutinize our predictive capabilities about their behavior with trials seems unnecessary (or sometimes completely ethically unacceptable). Defibrillation of a dysrhythmic heart as a treatment procedure was never put under any test of clinical trials but no one insists it must be trialed to use it as a medical intervention.[25]

In recent papers taking notice of such criticisms, Russo and Williamson have loosened their conditions about the necessity of having both the evidence of difference-making and mechanisms (especially in Williamson 2019). The relaxed version of RWT drops the descriptive part of the thesis but still imposes strong normative considerations. The relaxed version of RWT has some argumentative weight. According to Williamson, medical professionals are far more ready to assess any causal claim if they have both types of evidence and RWT, should in such cases be considered as a complete medical causal epistemology. At first, this does seem to be a reasonable statement. Associations or correlations are population level relations. But they do not arise out of nowhere: no epidemiologist would deny this. They are grounded in the workings of underlying biological mechanisms. Yet our knowledge of mechanisms which supposedly ground associations is often partial. Even if we have almost complete descriptions of mechanisms, we cannot theoretically disregard the option that at least some of them behave stochastically and, as I have showed, that multiple mechanisms can and often do lead to the same outcome. Therefore, even if we know a lot about how some mechanism is constituted and how it works, we search for the difference-making evidence for at least three reasons: (i) to corroborate that this mechanism is indeed capable of producing the outcome, (ii) to have information about how much of the outcome varies due to the mechanism in question, and (iii) to find out how much of the outcome in the population under study is indeed produced by the given mechanism. Consider again scientists trying to understand the

[25] The mantra of EBM-ers "if it wasn't clinically trialed, it isn't evidence" was mocked in, for example "Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomized controlled trials" by Smith and Pell (2003) and "Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial" by Yeh et al. (2018).

pathogenic processes induced by cigarette smoke. To understand how, in general, carcinogenic chemicals in cigarette smoke cause lung cancer only (i) seems to be informative. Once we have a mechanistic explanation connecting smoking and lung cancer we would like to know (ii) and (iii), but they are not necessary for such an explanation. Even if such a mechanism is indeed capable of producing the phenomenon, it does not need to do it in any of the cases of lung cancer, no matter how far-fetched such a case would be. Therefore, (ii) and (iii) are evidence of population level causal relations and they tell us *that*, rather than *how*, a mechanism produces a phenomenon to the degree it does in a population.

1.6. Mechanistic reasoning and medical treatments

The history of medicine is full of examples of treatments based on pathophysiological rationale, but which had, in the end, negative effects or even terrible consequences. One does not need to try hard to find these examples. Bloodletting and leaches were used as medical treatments throughout the 19th century and justified by mechanistic reasoning and mechanistic explanations of phenomena. Mercury was used in the form of ointments to treat syphilis, but it caused numerous patients to develop serious metal poisoning with often lethal consequences.[26] As mentioned before, the lack of an identifiable mechanism linking cause and effect, in some cases delayed an effective treatment or preventive strategy (for example, in Semmelweis's case). Failed predictions of medical interventions based on mechanistic reasoning continued long into 20th century. In his (2011a), Howick compiled numerous examples of failed mechanistic reasoning in the 20th century medicine. These examples show (i) how causal claims based on mechanistic evidence suggested positive outcomes of treatments while clinical trials revealed them as either ineffective or harmful, and (ii) cases where mechanistic reasoning delayed the acceptance of treatment based on the population trials because the mechanism was not yet identified.

Howick specifically discusses the use of antiarrhythmic drugs throughout the 1970s and 1980s for the treatment of arrhythmias in patients who had survived a heart attack. At the time, the mechanisms involved in the pathophysiology of arrhythmias and behind the development of antiarrhythmic drugs were thought to be well understood and these drugs were put into use

[26] The expression "One night with Venus, a lifetime with Mercury" comes from the use of mercury compounds to treat different venereal diseases, especially syphilis.

before any serious clinical trial had been performed. Later clinical epidemiological research revealed that these drugs increased rather than decreased mortality. Howick in his (2011a) quotes several studies which estimated that, worldwide, these drugs have killed more people than were killed in action during the whole Vietnam war. In her (2012), Andersen discusses an example of the administration of prophylactic paracetamol with infant vaccination. The mechanistic reasoning behind the administration of both treatments simultaneously indicated that these mechanisms should not interfere, that is, there were no known causal pathways that these mechanisms could share. However, the trials suggested that these mechanisms indeed interfered and resulted in compromised disease immunity.

Biological hypotheses are constantly changing, improving, or being abandoned. What holds today will not necessarily hold tomorrow. The take-home, as often stated, is that we should not make predictions about the efficacy of medical treatments based on biological knowledge – mechanistic reasoning. This constituted much of the rationale governing the emergence of EBM at the beginning of the 1990s and the rise of its popularity until, eventually, it became the primary medical evidential framework in the 21st century.

However, not all failures of mechanistic reasoning led to undesirable and tragic consequences. Some of them were in fact quite serendipitous discoveries. I present in this section the case of the administration of sildenafil citrate to treat erectile dysfunction. The compound was developed as an intervention into the NO-cGMP causal pathway to treat angina pectoris and hypertension. Although it failed to produce desirable effects in treating angina pectoris and hypertension, it was discovered to have other beneficial effects, and remains one of the most financially beneficial pharmaceutical discoveries to this day.

Here is a brief history of the research on the NO-cGMP pathway and the intervention into the pathway by the compound sildenafil citrate.

Immediately following the discovery of cyclic adenosine monophosphate (cAMP) in 1958 scientists presumed that there could also be other cyclic nucleotides involved in different cellular regulatory processes and pathways. Cyclic guanosine monophosphate (cGMP) was first synthesized in 1960, and in the following years the existence of endogenously produced cGMP was confirmed from rabbit urine (Kots et al. 2009). These studies also indicated that cGMP could be degraded by the same type of enzymes responsible for the degradation of cAMP. In the next 10 to 15 years, researchers discovered that cGMP levels are regulated by

the enzymes guanylyl cyclase and phosphodiesterase. By the 1980s, scientists researching the behavior and properties of smooth muscle cells and endothelial cells in bringing about smooth muscle relaxation had identified several molecules that figure in this mechanism (such as soluble guanylate cyclase (sGC), cGMP, and different types of phosphodiesterase).

In the meantime, a group of researchers centered around Robert Furchgott found that vasodilation, that is the relaxation of blood vessels, is in some way dependent on a factor derived from endothelium. But what that factor was remained unknown. Hence, it was labeled as endothelium-derived relaxing factor (EDRF). In the meantime, Ferid Murad and his group had found that nitroglycerine activates an enzyme important for the synthetization of cGMP. Simultaneously, Murad and Louis Ignarro suspected that EDRF could be a nitrate and that this factor increases the cGMP synthesis. Finally, both Murad and Ignarro independently identified that EDRF was, in fact, nitric oxide (NO). This discovery was quite surprising since it was not suspected that NO could be produced endogenously.

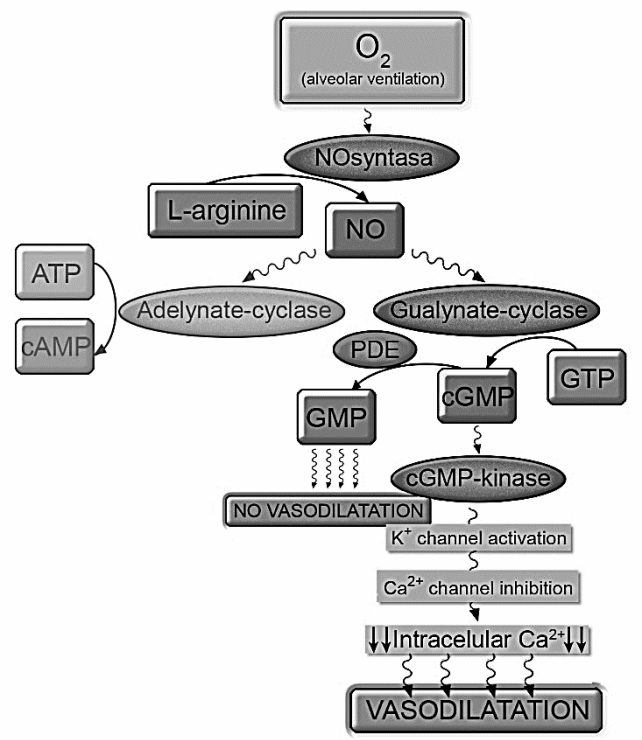


Figure 6. The NO - cGMP causal pathway.[27]

[27] By BQmUB2012010 - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=22800790

The causal pathway issuing in the relaxation of smooth muscle cells and, therefore, vasodilation, was now complete. NO is formed from amino acid L-arginine by nitric oxide synthase. In the corpus cavernosum (the sponge-like erectile tissue in the penis), NO is formed in the cavernosal nerves and then diffuses along the erectile tissue. It activates soluble guanylyl cyclase which in turn forms cGMP. cGMP in turn binds to and activates protein kinase G (PKG). This activation of PKG results in a number of chemical reactions leading to a decrease of levels of calcium which eventually result in smooth muscle cell relaxation. Finally, as hinted above phosphodiesterase (PDEs) were identified as a group of enzymes which can activate the degradation of cGMP if its levels are too high. In pathological states, however, where cGMP levels are below ordinary, and the degradation of cGMP by PDEs has negative effects since blood vessels cannot widen to allow for better blood flow.

However, the admission of nitrates in order to raise the level of cGMP proved ineffective because of the decrease in response after prolonged administration. Scientists, therefore, had to look for an alternative approach, one which would avoid this problem. Soon afterwards, a PDE enzyme was recognized as an optimal target. Since PDE5 enzyme specifically degrades cGMP, inhibiting its action would allow for the dilation of blood vessels. This is not the case for other PDEs, since, for example, PDE1 and PDE2 degrade both cAMP and cGMP. Following this rationale, PDE5 was chosen as the target of intervention. The development of a compound which could bind to PDE5 and prevent it to degrade cGMP shortly followed the above presented discoveries of mechanisms and molecules involved in the physiology of the cardiovascular system and the pathology of hypertension and different cardiovascular diseases. Finally, the chemical compound sildenafil-citrate (UK 92,480) was first synthesized in Sandwich laboratories of Pfizer in United Kingdom in 1989 (Ghofrani et al. 2006). Preclinical trials showed promising results, and the compound finally entered clinical trials in 1991.

In these first clinical studies, from 1991 to 1992, researchers found that sildenafil citrate had limited results in lowering blood pressure. In addition, it interacted with nitrates, which were usually administrated to patients with angina, and this led to a noticeable decrease of blood pressure in some cases (Osterloh 2004). Furthermore, admission of sildenafil citrate had another effect, which, at the time, was considered as an adverse effect – penile erection. Because of the study population containing young healthy males and having limited predictability of its adverse effects in middle aged population with comorbidities such as

diabetes, researchers were wary of exploring this potential new use of the compound (ibid). Nonetheless, the penile erection effect of sildenafil citrate was something worthy of investigating further. By late 1993, a new set of clinical studies was undertaken, but now with the new outcome of interest. Finally, by 1998 Pfizer applied for FDA and EMEA (European Medicine Evaluation Agency) registration of a new drug, named as *Viagra*, and both agencies approved it in the same year (Ghofrani 2006). To this day, *Viagra* remains one of the most sold and prescribed drugs in the world. Although very sound at the moment, positive patient outcomes of *Viagra* were discovered only after a failed instance of mechanistic reasoning.

Certainly, contemporary medicine values the knowledge of biological mechanisms. Most of our treatments have come and still come from the investigations of basic sciences. But, on the other hand, contemporary medicine disvalues its predictive power. In the medical literature, this claim is a conclusion of enumerative induction – mechanistic reasoning has failed numerous times in the past, and most likely, it will continue to do so. In the rest of this dissertation, I will offer a philosophical analysis of mechanisms in medicine and mechanistic reasoning in order to answer the following questions: (i) what is a mechanism, (ii) what is a mechanistic explanation of a phenomenon, (iii) what, exactly, does mechanistic reasoning amount to, (iv) why does it fail so often, and, finally, (v) how and when can mechanistic reasoning be of a good quality.

2. MECHANISMS IN MEDICINE: EXPLANATION

Abstract

In this chapter, I discuss mechanistic philosophy and its core commitments in the context of explanations of biological and biomedical phenomena. The chapter is structured as follows: after presenting a general introduction to the ideas and arguments that influenced the rise of mechanistic philosophy in 1990s, I argue that mechanistic philosophy is constituted of three theses – an ontological, an epistemological, and methodological thesis. Next, I discuss what those theses amount to in the medical sciences and practice. In the final two sections of the chapter, I propose and discuss my view on the following issues: (i) the relation between ontological mechanisms (the supposed real mechanisms in nature) and their representations (models of mechanisms); (ii) the criteria for a good mechanistic explanation, and finally; (iii) how are diseases qua dysfunctions explained within the mechanistic framework.

2.1. Introduction: "The New Mechanistic Philosophy"

The concept of mechanism has been discussed in philosophy since at least the 17th century. The mechanistic view of nature was influenced directly by the incredible achievements of scientists like Descartes, Boyle, and Newton. The view of mechanical interactions between bodies as the cause of natural phenomena became the principal view of nature itself – nature was mechanized. Since then, mechanisms have been sporadically discussed in philosophy. For example, Simon (1962), Kauffman (1971) and Wimsatt (1976) were the forerunners of the modern mechanistic philosophy and most of the core ideas of the modern mechanistic philosophy were already introduced in those papers. For example, Wimsatt writes: "At least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms, rather than explaining theories by deriving them from or reducing them to other theories, and this is seen by them as reduction, or as integrally tied to it" (Wimsatt 1976: 671). However, not until the early 1990s had mechanisms entered the focus of mainstream philosophy of science. There are several important lines of thought that influenced the rise of mechanistic philosophy in the 1990s: Salmon's ontic conception of explanation (1984), the shift of focus of philosophy of science from physics to special sciences, announced in papers by, for example, Herbert Simon (1962), Stuart Kauffman (1971), and William Wimsatt (1972, 1976), arguments against the regularity theory of causation and the logical positivist account of laws (Cartwright 1983, Salmon 1984), and arguments for the autonomy of the special sciences (for example, Kitcher 1984).

By the 1990s and early 2000s philosophers of science (e.g., Bechtel and Richardson 1993/2010, Glennan 1996, 2002, Machamer, Darden and Craver 2000 (hereafter MDC 2000), and Bechtel & Abrahamsen 2005) started arguing that causal explanations in the life sciences predominantly use the concept of mechanism rather than law of nature. In their often-quoted paper, Bechtel and Abrahamsen begin by noting that the notion of "law" in scientific papers from different fields of biology is almost completely absent when explanations of phenomena are considered, and rarely if ever mentioned in any other context (e.g., prediction). But, as these authors claim, looking closely at the literature in biology and its subfields, a rather different concept emerges. They write: "Perusing the biological literature, it quickly becomes clear that the term biologists most frequently invoke in explanatory contexts is mechanism" (Bechtel and Abrahamsen 2005: 422). Some have labeled the shift in focus of philosophy of science to mechanisms and causes rather than laws of nature as the methodological turn in the philosophy

of science since it changed the way philosophy of science was conducted and it perhaps changed our perspectives of what the goals of the discipline itself ought to be. In that regard, H.K. Chao et al. write: "That means what really matters to philosophers of science, and what philosophical discussions should be based on, is what scientists actually do and how they do it rather than philosophers' visage of what science is and how scientists should do it" (2013: 1,2). It is this bottom-up methodology that so nicely characterizes the new mechanistic turn in philosophy of science. Rather than discussing what the notions and concepts from science mean and how scientists should shape their research and explanations on the basis of prior philosophical clarifications, mechanistic philosophers turned their attention to scientific practice in order to shape their own line of research. In his foreword to the latest and the most comprehensive volume on mechanistic philosophy, Wimsatt states that this methodological turn means "to take the work and claims of scientists seriously, and to look at what they can bring to philosophy rather than to suppose that the primary mission of philosophers is to bring edification to scientists" (Wimsatt 2018: xvi).

In that regard, biologists, according to mechanistic philosophers, describe causal relations leading to or producing a phenomenon rather than deducing the phenomenon from initial conditions and laws of nature. Craver formulates this in the following way: "In an explanatory text, the *explanandum* is a description of the phenomenon and the *explanans* is a description, or schema, of a mechanism" (Craver 2007a:139). Biologists, then, explain a phenomenon by describing it as a product of a certain biological mechanism or as if a phenomenon itself is a mechanism (that is, rather than producing a phenomenon, a mechanism is constitutive of a phenomenon). The explanation of a biological phenomenon is not a deductive argument where the explanandum is deductively inferred from the premises and laws of nature, nor it is a matter of providing a unifying account of a range of different phenomena. Rather, is it a description of a causal structure – a mechanism – responsible for the phenomenon. The phenomenon is explained by providing an account of the inner workings of such a mechanism and of how the mechanism's parts and its overall organization bring about the phenomenon.[28]

Since the 2000s, however, mechanisms have been discussed in different areas of philosophy. They have been discussed in philosophical accounts of causation, explanation,

[28] I will present and discuss mechanism's function and phenomena in more detail in section 2.5.2. For now, consider a phenomenon as whatever the end product of a certain mechanism is.

prediction, confirmation, and in the realism/antirealism debate in philosophy of science. The widespread appeal and reference to mechanisms across philosophical discussions, as well as the high convergence of ideas on mechanisms in these discussions, has led some to consider that, by now, we can claim that there is a distinct and consistent philosophical view, ranging from philosophy of science to metaphysics and epistemology – "The New Mechanistic Philosophy".[29] For example, at the beginning of the third chapter of his book from 2017 Glennan argues that the philosophy of the "New Mechanicism" is not just an account of a particular kind of explanation found in biological sciences but it has rather evolved into a general philosophical position or a worldview on what and how the world is and how the sciences try to grasp it and explain it. In that regard, Glennan writes, "The New Mechanical Philosophy is both a philosophy of nature and a philosophy of science. It tells us something about how the world is, as well as something about how we, particularly through the methods and institutions of science, may come to know that world" (Glennan 2017: 59). But the concept of a mechanism from "The New Mechanical Philosophy" diverges from the one that Salmon used. Glennan (2002) is perhaps most explicit in his view of the difference. He writes: "Salmon/Railton mechanisms are sequences of interconnected events while complex-systems mechanisms are things (or objects)" (Glennan 2002: S345). Salmon's mechanisms are processes while Glennan's mechanisms are things; Salmon's causal processes expand into a causal network, while Glennan's mechanisms are like "chunks", in some way divided from their environment, and with sometimes more and sometimes less apparent boundaries.

So, what are the mechanisms of "The New Mechanistic Philosophy"? Let us consider the three most-cited definitions of mechanisms from the philosophical literature. Although different in some respects, these three definitions of a mechanism have become a somewhat canonical in mechanistic philosophy. They have enough in common to unpack three main features or characteristics of mechanisms that all mechanists can agree upon.[30]

> Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.
>
> Machamer, Darden, and Craver 2000, 3

[29] The term first appeared in an article by Skipper and Millstein (2005).

[30] For present purposes I will only present these preliminary observations of mechanisms and mechanistic philosophy. I will discuss them in a more detailed manner in the following sections.

> A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.
>
> Glennan 2002a, S344

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena.
>
> Bechtel and Abrahamsen 2005

The first feature that all mechanists agree upon is that mechanisms are presumed to be complex causal structures in the sense that they are "compounds", which are "composed of simpler things" (Glennan 2017: 19).[31] Usually two kinds of components are thought to constitute any mechanism: component parts (a finite number of constituent entities or working parts) and component activities (or interactions between parts). The second feature of mechanisms (and the most distinctive contribution of mechanistic philosophy) is that their parts and activities (and/or interactions) are organized so that they (regularly) produce a phenomenon. The organization of parts and their causal relations is what makes these systems distinctively "mechanistic". It is the organization that separates mechanistic systems from aggregative systems. If either parts or activities were organized differently, then they would no longer produce the same effect. A different organization leads to either no production at all or the production of different effects. For example, consider an airplane. There is some specific organization of its component parts and their interdependent causal relations that all together make the airplane capable of flying. There can only be minor variations in this organization. Usually, organization includes spatial, temporal, and causal organization: among other things, these include the distributions of parts within the mechanism, temporal orders of their causal relations, and different signaling relations such as feedback and feedforward signals. It is

[31] What is the difference between systems and mechanisms? Illari and Williamson (2012) argue that complex systems imply a rigid organization while mechanisms, as they conceive them, are more flexible and open to adjustments as they continue to work. A system is composed of different mechanisms, that is, mechanisms are instantiated in systems. For example, the circulatory system in mammals is constituted of numerous mechanisms (the mechanism for vasodilation and vasoconstriction being just one of many). The majority of (or perhaps all) mechanistic philosophers accept this particular relation between systems and mechanisms when those two are considered together.

because parts and their activities and interactions are specifically organized that mechanisms exhibit "orchestrated functioning".

The third feature is not constitutive of a mechanism, but it is equally important in its own right. Mechanisms are not mechanisms *simpliciter*. They are sometimes considered to "perform functions", and sometimes as "mechanisms for a behavior" (Kaufmann 1971, Glennan 1996, 2002, Craver 2007a, Garson 2013).[32] A heart is a complex system *for* pumping blood. An airplane is a mechanism *for* flying.[33] So mechanisms ought to underlie regularities. However, this is not the only way mechanisms have been conceived. Ioannidis and Psillos (2018) recognize that there are two different ways mechanisms have been discussed so far: "mechanisms-for" which correspond to the causal structures underlying regular behavior, and "mechanisms-of" which correspond to Salmon's causal-mechanical explanations and Glennan's ephemeral mechanisms (2010, 2017).[34] Ioannidis and Psillos claim that "mechanisms-of" mechanisms are the underlying causes responsible for one-off events. It is just a causal process connecting cause and effect. Here, however, I will concentrate mostly on their "mechanisms-for" conception of mechanisms since it is this conception that mechanistic philosophers usually think of when they talk about mechanisms from the life sciences.

Many mechanists (especially Glennan and Illari and Williamson) argue that the concept of mechanism has a wider reach than being merely an explanatory framework of special sciences. For these mechanistic philosophers, "mechanism" is an ontological view of a world populated by mechanisms. Mechanisms, as these are described by mechanistic philosophy and science, underlie almost all of the phenomena in the world (excluding some phenomena from fundamental sciences): from volcanos as mechanisms for spewing lava to mechanisms underlying our Solar system or even natural selection. In order to account for all these intuitions

[32] Garson in (2018), however, argues that phenomena should be considered as being constitutive of mechanisms (in the same way that entities, activities, and organization are) and not something only incidental to them.

[33] Interestingly, some philosophers now call this *Glennan's law* since he was the first to emphasize it. According to Illari and Williamson (2012), the term *Glennan's law* was first used by Darden and Craver in a conference in Kent in 2009.

[34] "On the other hand, there are certainly processes involving entities that engage in activities and interactions that produce some phenomenon, but where that process is not systematic or repeatable. To the extent that these mechanical processes are not systematically organized, they are instances of what I call ephemeral mechanisms" (Glennan 2017: 27).

about mechanisms and mechanistic explanations, Illari and Williamson, followed by Glennan, propose a minimal conception or definition of mechanism:

MINIMAL MECHANISM:

> 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: 120

> A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon.
>
> Glennan 2017: 17

As said, this minimalist view of mechanisms is shaped by the ontological motivations of some mechanistic philosophers (Bechtel would probably be one of the mechanistic philosophers sitting on the opposite side of the ontological table in this debate). Does the notion of mechanism then become vacuous and over-permissive? This is certainly an interesting question, but I will leave it for the later sections of this chapter. The problem, I believe, emerges if one attempts to give a rich ontological framework with the concept of mechanism at its foundation rather than make claims about the epistemological and methodological commitments of the life sciences. However, unless explicitly stated, I will use these two definitions when discussing mechanisms as they are supposed to be found in nature. Illari and Williamson claim their definition is a consensus, and Glennan proposes his own although he acknowledges that it has been influenced by Illari and Williamson's definition and their argument for finding a consensus definition of mechanisms in the first place.

2.2. Three Theses of "The New Mechanistic Philosophy"

Mechanisms are used and discussed in different ways in different areas of philosophical research: as a type of causal structure or metaphysical entity (for example, as an artifact, a machine, or a biological structure); as certain configurations of entities individuated by their causal organization; or as a type of explanatory and methodological practice in sciences. Philosophers talk about mechanisms of neurotransmitter release (Craver 2007a). They argue for a mechanistic interpretation of natural selection (Barros 2008) or a mechanistic

interpretation of solar system eclipses (Illari and Williamson 2012). That is, mechanisms are discussed in almost all of the traditional topics of philosophy of science. Many philosophers have noticed then that there are different senses of the term mechanism used in these discussions and that we should make these senses explicit. For example, Woodward (2013) writes:

> An obvious worry raised by arguments over whether there are non-mechanistic forms of explanation, either in biology or elsewhere, or whether all biological systems are 'machine-like', is that such disputes are (or threaten to become) largely terminological or semantic – the answer depends on what is meant by 'mechanism', 'mechanical explanation', and so on, and these are notion without clear boundaries.
>
> Woodward 2013: 40

It is often covertly assumed in the mechanistic literature that there is a clearly understandable use of the term mechanism across the sciences. So, for example, when biologist talk about the mechanisms of protein synthesis, when economists talk about the market mechanisms, when geologists talk about the mechanisms of plate movements, and so on, they all refer to one and the same concept.[35] But contrary to what the prime motivation of the mechanistic methodological turn in philosophy of science is, Lenny Moss observes that this implication of mechanistic philosophy is not so straightforwardly revealed in the practice and theory of the different sciences. Contrary to a general view of mechanistic philosophy, Moss insightfully observes that "[t]here is no place in an undergraduate course curriculum where, for example, a student is instructed as to the ways in which the term 'mechanism' means something different in the context of a chemistry laboratory versus that of a biology laboratory, nor would a full grasp of this distinction be easily conveyed to a lay person" (Moss 2012: 165). There is simply no univocal or clearly understandable meaning of mechanism even within the same scientific field, let alone across different sciences.

Furthermore, it seems that scientists in biology and biomedicine often use other causal concepts such as systems, causal or biochemical pathways, cascades, triggers etc., and these sometimes coincide with the meaning of mechanism is in the philosophical literature while sometimes they, arguably, mean different things. Therefore, some authors have argued that the

[35] After all, this is what guides Illari and Williamson's proposal for a consensus view on mechanisms mentioned in the previous section.

notion of mechanism does not adequately capture all the peculiarities of these other notions as they are used by biological and biomedical scientists (Boniolo and Campaner 2018, Ross 2020). Ross, for example, claims that causal pathways (e.g., developmental, gene expression, metabolic, anatomical, or ecological pathways) differ from mechanisms. A causal pathway, as she argues, refers "to a sequence of causal steps that string together an upstream cause to a set of causal intermediates to some downstream outcome" (Ross 2020: 137). Ross argues that the concept of causal pathway tracks the flow of some entity or signal through a system. By doing so, it intentionally and significantly abstracts causal details and emphasizes "the 'connection' aspect of causal relationships" rather than the organization, the fundamental aspect of any mechanism (Ross 2020: 139). A single pathway, then, can be instantiated in different mechanisms (e.g., NO-cGMP pathway in both smooth muscle cells and in the retina). Craver and Darden (2013), Bechtel (2019), and Brzović, Balorda and Šustar (2021) have discussed the relation between pathways and mechanisms in terms of constructing mechanistic explanations, and the explanatory virtues and priorities of these concepts. Nonetheless, prior to these discussions, mechanistic philosophers do need to address Moss's argument that it is mechanistic philosophers' unfounded assumption that when that notion is used in the sciences it always refers to one and the same explanatory, methodological, or metaphysical thing. If one really wants to state what referring to "mechanism" means in different sciences, perhaps the most accurate meaning after all is that of a cause or a causal sequence. All other meanings, then, are added by the philosophical analyses.

In his (2012), Daniel Nicholson presents another notable attempt to resolve the apparent ambiguity of the notion of mechanism in philosophy. Nicholson identifies at least three different senses of the notion that are interchangeably used by philosophers. The first sense is what Nicholson labels as *Mechanicism*. He defines this as a philosophical thesis underlying mainstream biological sciences since at least 17th century. This is a thesis that "conceives living organisms as machines that can be completely explained in terms of the structure and interactions of their component parts" (Nicholson 2012: 153). The main desiderata of *Mechanicism* are the "ontological continuity between the living and non-living", the analogy to man-made machines and "the commitment to reductionism in the investigation and explanation of living systems" (Nicholson 2012: 153). The second sense is *Machine mechanism*. This is more of a methodological and explanatory stance taken by the proponents of *Mechanicism*. It is a stance used "to describe machine-like systems, or rather, systems conceived in mechanical terms; that is, as stable assemblies of interacting parts arranged in

such a way that their combined operation results in predetermined outcomes" (Nicholson 2012: 153). The third sense is *Causal mechanism*, where a phenomenon is explained by specifying a step-by-step sequences of causes that gives rise to it.[36] Causal mechanism in this sense resembles Salmon's ontic explanation of causal structures in the world: "Causal processes, causal interactions, and causal laws provide the mechanisms by which the world works; to understand why certain things happen, we need to see *how* they are produced by these mechanisms" (Salmon 1984a: 132).[37] It does not imply that there are specific kinds of causal structure which are "mechanisms" as distinct from, for example, the networks of some kind of difference-making relations like a network of counterfactual conditionals, a network of intersections of Salmon's causal processes, or something else.

Both Nicholson's and Moss's discussions of a concept of mechanism are interesting contributions to the debate on mechanisms. Both papers were published when "the mechanism craze" in philosophy was at its height and both papers failed to ignite a serious response from "the mechanists". Similarly, I argue that there are three different theses constituting "The New Mechanistic Philosophy". First, mechanisms refer to a productive, causal structure underlying phenomena, regularities, and functions. That is, mechanisms are real things or objects in the world. Second, mechanisms refer to an explanatory notion, a distinct type of scientific explanation found across the life sciences (as well as in other sciences). The third thesis is that mechanisms refer to a specific methodology of scientific inquiry. My distinction of mechanistic philosophy into these three theses is influenced and very similar to a distinction made by Levy in (2013). Levy also argues that there are three different theses of "mechanism" which can be interchangeably found in "The New Mechanistic Philosophy" under the same notion. However, rarely if ever are they clearly distinguished in the mechanistic literature. Although my view converges with Levy's, there are differences. Levy's first thesis and my first thesis are different, and although our second and third theses seem to be identical, I take my explication of the theses to be more detailed and slightly more relevant concerning the particular issues in the science and practice of medicine.

[36] Similar to Nicholson, Ionnaidis and Psilos (2018) claim that the mechanistic stance is nothing more nor less than providing an explanation of a phenomenon by citing its causes. They do not take that commitment to the mechanistic stance or mechanistic explanation necessarily invokes the first two Nicholson's senses of mechanism. I will present their view in more detail in section 2.5.1.

[37] It is because of this that Salmon's conception of a causal explanation which explains an event by locating it in the causal structure of the world is also called the causal-mechanical account of explanation.

The first thesis in Levy's account is the thesis of *Causal mechanism* (CM), and he sees it as a contribution to a discussion in metaphysics, particularly on causation, rather than in the philosophy of science. It is a thesis usually associated with Glennan's work (1996, 2002, 2009, 2017), and it states that all causal relations (excluding those in fundamental physics, if causation is found at all on that level) exist in virtue of underlying mechanisms. Williamson offers a similar definition of a mechanistic theory of causality: "A mechanistic theory of causality holds that [...] two events are causally connected if and only if they are connected by an underlying physical mechanism of the appropriate sort" (Williamson 2011: 421). Levy claims that Glennan offers a general theory of causation and as such it is a rival to other general theories of causation (e.g., the regularity theory or counterfactual theory). It could be claimed that, quite possibly, Glennan's account is simply the only real contender for a mechanistic theory of causation. Certainly, Glennan admits he intends to provide an analysis of causation for all of its purported cases, excluding fundamental science. However, arguments could be made that Machamer, Darden and Craver's (2000), and Machamer (2004) are accounts of causation *for* biological sciences; that is, they are theories of causation of limited scope and applicability.

Although CM is a thesis held by at least one philosopher I claim that there is another thesis of mechanistic philosophy similar to the CM thesis though very much distinct from it. The consequences of that thesis, however, I find to be far more relevant for the present discussion. As far as I know, the majority of mechanistic philosophers deny that a plausible general theory of causation can be developed through the concept of mechanism, yet they still accept the ontological reality of mechanisms. That is, mechanisms, as described by true and complete mechanistic explanations, exist *as such* in nature and therefore, mechanistic philosophy accurately describes the ontological furniture of our world. Where the CM thesis had only one advocate (that we can surely acknowledge), this thesis is present in almost all of the works associated with the most eminent mechanistic philosophers. Craver, among other mechanistic philosophers, is very explicit about this claim:

> There are mechanisms (the objective explanations) and there are their descriptions (explanatory texts). Objective explanations are not texts; they are full-bodied things. They are facts, not representations. They are the kinds of things that are discovered and described. There is no question of objective explanations being 'right' or 'wrong, or 'good' or 'bad'. They just are.
>
> Craver 2007a: 27

Mechanisms are out there. They produce phenomena. They can be studied and revealed. It is the business of the sciences to find them and understand them. Let me then replace CM with a thesis of the ontological reality of mechanisms – *Ontological Mechanicism*.

OM: Mechanisms that our sciences describe exist as real things in nature.

Scientific realism is a characteristic of mechanistic philosophy. Its influences in mechanistic philosophy can be traced back to Salmon's ontic conception of scientific explanation and the view that a causal connection between an exposure and an outcome requires identification of real and *actual* mechanism, not merely plausible or possible ones (or some abstract entity). For example, Illari and Williamson argue that explanation of a mechanism, if it is going to be explanatory at all, has to describe something real "out there". They write: "The mechanism itself is different from any model, schema or other description or representation of *that* mechanism, and the mechanism itself is real" (Illari and Williamson 2011: 827). Mechanistic explanations are explanatory because they describe or represent real mechanisms in nature. Whether causation can be understood through mechanisms or whether mechanisms require a further metaphysical account of causation is a different question.

Carl Craver is commonly referred to as the most well-known advocate of the strong reading of the ontic account of mechanistic explanation, while William Bechtel is usually considered as the best-known advocate of the epistemic account. The distinction between ontic and epistemic accounts of scientific explanations comes from Salmon's discussion of the differences between Hempel and Oppenheim's "covering law" account of explanation (CL) and Salmon's causal-mechanical account. In Salmon's conception of scientific explanation, causal relations in the world just are the explanans, while the phenomena, which these causal relations are responsible for, are the explananda. Causal relations explain the effect. Since an explanation can be explanatory only if it reveals the underlying causal structure of our world, Salmon takes such an explanation as an ontic conception of explanation. On the other hand, in the CL account, an explanation is informative because of its logical structure and hence, for Salmon, it is an epistemic kind of explanation.

Mechanistic philosophers have since applied the ontic/epistemic distinction and terminology to the differing views on what exactly is explanatory in the mechanistic account of explanation. That is, in this debate, the issue has been structured around the problem of what does the explaining: the relations in the world that mechanistic explanation reveals or the

epistemic features of the mechanistic explanation itself? For Craver, mechanistic explanations are explanatory because they explain accurately and in as much detail as possible the real, ontological mechanisms, while for Bechtel, mechanistic explanations, although explaining real mechanisms, are not identical to the mechanisms that they explain. Explanations of mechanisms add certain epistemic features which ontological mechanisms do not possess. These features – such as abstractions and idealizations – are added for the purpose of our understanding the phenomenon. Without these, our understanding of how a mechanism works is out of reach or is at least partial and insufficient. Ontological mechanisms, as advocates of the epistemic conception of explanation claim, do not explain anything by themselves (i.e., as they are in nature). Explanation is an epistemic category, not ontological or metaphysical. If mechanistic explanation is to be ontic, as advocates of the ontic account of explanation claim, the things it describes have to exist *as such* in nature. Following such a view, mechanistic explanation becomes better and more informative when it describes its target mechanism more accurately and in more detail; or so it is presumed.

Nonetheless, if we set this debate aside for a moment, most if not all advocates of the epistemic approach to mechanistic explanation share the sentiment that what is described are in fact the real mechanisms in nature. Bechtel, as the most well-known advocate of the epistemic approach, writes in an article coauthored with Abrahamsen: "Our own approach is to begin with a basic characterization of mechanisms *as found in nature* and then (see below) elaborate it into a framework for mechanistic explanation" (Bechtel and Abrahamsen 2005: 423, emphasis added). Later in the text, they make further explicit ontological claims about mechanisms: "mechanisms are real systems in nature, and hence one does not have to face questions comparable to those faced by nomological accounts of explanation about the ontological status of laws" (Bechtel and Abrahamsen 2005: 424, 425). Whether mechanisms as they are described by mechanistic explanations exist in nature is far more important question when assessing the further consequences of mechanistic philosophy than whether all causal relations, excluding ones from the fundamental level, can be reduced to mechanisms.

This leads us to the second thesis of "The New Mechanistic Philosophy" - *Epistemic mechanicism* (**EM**). EM is a thesis about the structure of scientific explanation, at least in the life sciences. It means that scientists approach biomedical phenomena with a certain explanatory framework that must be satisfied. For example, Darden writes: "When the goal is to find what produces the phenomenon, then one searches for a mechanistic type of hypothesis"

(Darden 2018: 256). EM is a thesis about what a good scientific (mechanistic) explanation in the life sciences should. In other words, EM is both a descriptive and normative account of scientific explanation. It possesses several metaphysical presumptions (about entities, their causal relations, and part-whole relations, as discussed especially in Craver 2007a) but its goal is not an account of causation or an ontological account of mechanisms. It is a thesis about a scientific explanation. Thus, the EM thesis can be defined as follows:

EM: Causal explanations in biological sciences are mechanistic explanations.

Levy recognizes two important aspects shared by the majority if not all EM accounts (e.g., MDC 2000, Glennan 2002, Bechtel and Abrahamsen 2005, Craver 2007a): components and organization. To explain something mechanistically, explanation must consist of a description of entities and their properties, and how those entities in virtue of their properties mutually interact. It must describe the kinds of activities these entities perform, and how their interactions are instantiated. Finally, mechanistic explanation has to describe how those entities and their activities and interactions are organized to produce just that effect. Organization is the final and the most distinctive aspect of such an explanation. It represents all the information about the spatial and temporal distribution of entities and their activities and interactions.

How do we achieve this? While the deductive-nomological model of the CL account of explanation is a deductive argument expressed propositionally, mechanistic explanation can take various forms. Consider the many ways you can explain or represent how a heart works, or how the internal combustion engine works, or how protein synthesis works, and so on and so forth. Of course, one of the most straightforward ways is by providing a textual description of how one stage leads to another until, finally, the phenomenon (pumping blood, the movement of the pistons, synthesis of a protein) is brought about. But textual explanations are not the only way to represent or explain mechanisms. Open any textbook on molecular biology or pathology and you will see diagrams, pictures, graphs which include boxes, arrows, and similar visual tools, all put in service of a better understanding of how mechanisms work. All these epistemic tools serve to one end: a good mechanistic explanation describes how a stipulated mechanism produces "regular changes from start or set-up to finish or termination conditions" (MDC 2000: 3), or how it is "performing a function" (Bechtel and Abrahamsen 2005: 423). Mechanistic explanation "constructs" the model of a real mechanism by following this rationale. EM is therefore an account of explanation which stipulates what sort of things

or features are explanatorily relevant, what things or features are redundant for this particular kind of explanation, and what is the "representational vehicle" of a mechanistic explanation.

Finally, the third thesis of "The New Mechanistic Philosophy" is about the methodology of scientific investigation. It is a descriptive and normative position about strategies used to construct a mechanistic explanation. Levy makes a similar observation and calls this use of "mechanism" *Strategic mechanisms*. Although all of the advocates of EM hold SM, these are still distinct theses, and the distinction is usually not well, or even at all, demarcated and discussed in the literature. If EM tells us what a mechanistic explanation should contain and represent, then SM tells us how to achieve that. In that regard, it is a claim about a "characteristically mechanistic style of doing science, which involves particular methods of representation, reasoning and understanding" (Levy 2013: 107). What is a characteristically mechanistic style of doing science? It means that scientists in life sciences often use methodology that resembles the dismantling of a mechanism in order to understand it. Levy, as we will see in the section dealing with this mechanistic thesis, does not discuss in detail how this methodological thesis diverges from the epistemological one. In that respect, I will call this thesis *Methodological Mechanicism* to differentiate my view from Levy's:

> **MM:** There is a specific mechanistic methodology used in discoveries and in constructing explanations of phenomena in the life sciences.

As with the EM thesis, MM is built upon several propositions. Probably the most important, but also the most contentious, is that of modular assembly. This is one of the most important methodological assumptions in the majority of scientific experiments. However, whether or not a mechanism of interest is indeed modular is a matter of empirical research and not a theoretical precondition. Nevertheless, the majority of investigations into mechanisms start with such an assumption. The modularity of mechanisms then leads to the most common and widely used heuristics of mechanistic investigative strategy: *decomposition* and *localization*. These were first elaborated and presented by Bechtel and Richardson in their (1993). In a nutshell, decomposition refers to the *tracking down* and the *breaking down* of a system in its constitutive parts (i.e., entities). It is a top-down view of explaining a phenomenon.[38] Localization, on the other hand, refers to the assignment of causal roles to these parts. Although a vital part of "The New Mechanistic Philosophy", the MM thesis tends to be

[38] Ross (2020) calls this "the drilling down" strategy.

less elaborated than EM. Besides Bechtel and Richardson's book length analysis of the aforementioned mechanistic strategies, notable attempts at the analysis of mechanistic investigative strategies can be found in Darden (2006) and Craver and Darden (2013).[39]

In the following sections, I discuss the OM, EM, and MM theses in more detail. I will argue that disambiguating the notion of mechanism in "The New Mechanistic Philosophy" is important for several reasons. First of all, it is important to show that mechanistic philosophy has spread into different branches of philosophy. Second, I will discuss the metaphysics of mechanisms, especially of causal relations between component parts to show how the mechanistic and epidemiological approaches do not necessarily have a different view on the metaphysics of causation. Third, I place particular focus on the distinction between the sense of mechanisms from the OM and EM theses. I will show how the failure to distinguish between the two senses of mechanism often generate problems which can be avoided by its acknowledgement. Fourth, I will show how the OM and EM (as well as MM but less importantly) can be held and discussed separately and how this reflects on the scope and reach of mechanistic explanation of biomedical phenomena.

2.3. Ontological mechanicism

What are mechanisms of the OM thesis supposed to be? I have already mentioned that Glennan in one of his earlier papers considers them as things or chunks distinguished from their environment by their relatively stable organization. Whether other mechanistic philosophers agree that mechanisms are things as Glennan conceives them is not straightforwardly clear. Nonetheless, all mechanistic philosophers seem to agree that mechanisms are causal structures that are distinct from their environment and located in a particular region of space and time. Mechanisms, then, have or at least ought to have boundaries by which we are able to identify and define them. However, how those boundaries are identified and whether they are there independently of the identifier is a point of discussion. Also, it is not a necessary feature of ontological mechanisms that they have a start or setup and a finish or termination stage. These stages usually reflect our interests and oftentimes they are just a kind of heuristics utilized to grasp the functioning of a mechanism. Mechanisms in nature can

[39] Though these themes are already announced and to a lesser degree presented in the MDC paper from 2000.

operate in cycles or loops where there is no real sense in which one stage can be designated as a starting point while some other stage is a finishing point. For such mechanisms, separation into stages is a matter of trying to make them understandable, that is, it is a feature of a representation rather than of a real mechanism. Recall from section 2.1. the three definitions of a mechanism in philosophy (Machamer, Darden and Craver 2000, Glennan 2002, and Bechtel and Abrahamsen 2005). Although different in some respects, all three accounts define a mechanism as a complex causal structure constituted of entities or component parts, their activities or interactions between entities, and their overall organization. Let me now present the main metaphysical considerations pertaining to these three features of mechanisms.[40]

2.3.1. Entities

The nature of entities is a metaphysical discussion par excellence. It is a vital part of an ontological account of mechanisms but certainly not an issue that arises solely from the mechanistic philosophy. It is far from the goals of this dissertation to engage in a discussion of the ontology of objects or processes. I will, however, present what kind of a thing a part of mechanism ought to be if it is to be considered a part of an ontological mechanism.

We can start with Glennan's remark that "the category of things New Mechanists have referred to as entities coincides with common conceptions of substance, at least those conceptions that countenance compounds as true substances" (Glennan 2017: 49, 50). If mechanisms are real complex causal structures in the world, then they ought to be composed of real things. Parts of a mechanism and causal interactions between those parts constitute Salmon's "causal structure of the world". Craver distinguishes explanations of actual mechanisms from mere possible or plausible models of mechanisms by claiming that models of actual mechanisms include parts which are known to be real and independent of any role they have within the representation of mechanisms. As he states, a representation of a possible mechanism "contains black boxes or filler term that cannot be completed with known parts or activities" (Craver 2007a: 129, 130). Glennan also stressed this point in one of his earlier

[40] An interesting question is whether we should include the functions from Glennan's law as a part of the OM thesis. As we will see, I discuss the functions of mechanisms and the functions of a mechanism's parts withing the context of the methodological thesis since many mechanistic philosophers consider functions in and of mechanisms to be a matter of scientific research interest rather than coming from the very nature of mechanisms.

papers: “The parts of mechanisms must have a kind of robustness and reality apart from their place within that mechanism” (Glennan 1996: 53). In his argument against the mechanistic interpretation of computational models, Weiskopf stipulates similar constraints by introducing the principle of a “*Real Components Constraint* (RCC) on mechanistic models” (Weiskopf 2011: 320). Similarly, Craver claims that entities must be “stable bearers of causal powers” (Craver 2007a: 131), while Kaiser calls them “*material objects* i.e. continuants” (Kaiser 2018: 119). Nevertheless, this does not mean that entities within a mechanism cannot be dispersed in spacetime. We can think of certain groups of entities within a mechanism as being stable bearers of causal powers, while also themselves being decomposable into parts, spatiotemporally distributed and perhaps spatiotemporally overlapping with other entities. I will show later in the text how and where can this take place.

Craver’s filler terms in the representations of mechanisms are supposed to be gradually replaced with entities that are real material objects, located in space and time. The latter are spatially extended. They can (and arguably must) be individuated by their particular shapes and sizes. For example, proteins consist of one or more chains of amino acid residues, that can fold in at least four different motifs. While entities might retain their shape or size while being parts of a mechanism, they might also change, depending on their function within the mechanism or the type of behavior they exhibit *as parts* of the mechanism. For example, Illari and Williamson write, “Some entities remain comparatively unchanged over time, but others are more transient, such as the mRNA that is made from DNA, used as a template to make a protein, and then broken down again straight away” (Illari and Williamson 2012: 129). The majority of mechanistic philosophers argue that entities are things which engage in activities and/or interactions by being the bearers of causal powers. For example, MDC write: “The neurotransmitter and receptor, two entities, bind, an activity, by virtue of their structural properties and charge distributions” (MDC 2000: 3). Being shaped or folded into a certain motif allows a particular kind of protein to bind to a particular operator on a gene. However, that does not mean that entities always need to manifest or express their causal powers or engage in activities/interactions in order to be parts of a mechanism. An entity can perform a function in a mechanism by occasionally engaging in some kind of activity or interaction with other entities, yet it does not have to be doing so all the time. Proteins of a certain kind do not need to bind to operators constantly in order to be working parts of the gene expression

mechanism.[41] The temporal organization or sequential activities of particular entities in a mechanism is often crucial for its "orchestrated functioning" - to borrow Bechtel and Abrahamsen's term. However, an entity, if it is a working part of a mechanism, has to engage in at least some activity or interaction. It cannot be a working part of the mechanism if it does not do anything throughout the mechanism's functioning, from start to finish, or from input stage to output stage.

2.3.2. Activities and interactions

What is the nature of causal relations between entities? Such ontological and metaphysical discussions on the nature of mechanisms were prominent in mechanistic philosophy in the late 1990s and early 2000s but have declined over the years. This issue is quite rarely discussed in recent publications. When mechanistic philosophers did discuss causation within mechanisms, three general views were usually defended, and then defined in a more detailed manner: the causal relations in which entities engage in are activities, interactions, or a combination of the two.

Activities taken at face value lead to probably the most controversial ontic account of mechanisms – MDC's dualist ontology of mechanisms. An activity is always an activity of some entity, but it is activities rather than entities that are "the producers of change" (MDC 2000:4). MDC's account of activity is influenced by Anscombe's analysis of causation (1970, 1993). Anscombe claims that we acquire the meaning of general causal notions, such as production or bringing about, only through their more specific instances like scrape, push, wet, carry, eat, burn, knock over etc. Without acquiring meanings of these specific causal terms we could not have acquired the meaning of the general concept of cause. If there is any common feature of every causal relation it is that the effects "...derive from, arise out of, come of, their

[41] Gene expression mechanisms are an especially interesting example since such mechanisms arise the question whether an entity needs to be an individual or whether we can take populations of things as working parts of mechanisms, for example the populations of some particular kind of protein (Anić 2021). Similar questions have arisen in the discussion of whether natural selection can be thought of and explained by a mechanistic account. Certainly, a representation or model of mechanism, as I will discuss later, can take populations or dispersed entities as working parts of a mechanism, but this does not imply that, metaphysically or ontologically, it is a single working part rather than many working parts of a mechanism.

causes" (Anscombe 1993: 91-92). Causation, therefore, should be analyzed only in its specific instances.[42] Activities such as bonding, pushing, attracting, or repulsing, play a central explanatory role in any mechanistic explanation. Just as parts or entities are characterized by their location, structure, and orientation in the mechanism, the activities will be characterized by their temporal order, rate, and duration (see MDC 2000). The hierarchical decomposition of mechanisms, according to MDC, stops at the level where some of the components (whether entities or activities) are taken as fundamental in some scientific field. Considering molecular biology and molecular neurobiology, MDC categorize their "bottom out activities" into four types: geometrico-mechanical, electro-chemical, energetic and electromagnetic. Further individuation conditions for activities are their mode of operation, directionality, polarity, energy requirements and the range of activity (MDC 2003: 5). MDC argue that their dualist position has a couple of important advantages over, what they call, the "monist" and "substantivalist" positions. First, accepting the metaphysical reality of activities reflects scientific practice where scientists investigate, as their loci of inquiry, the specific activities of entities such as hydrogen bonding or membrane depolarization. Second, some properties such as charge can only be identified when they are manifesting themselves, i.e., when the entities that possess them engage in activities.

Entities, as noted above, engage in different activities. The same entities can engage in different activities, while different entities can engage in same activities. What distinguishes one mechanism from another or as being a mechanism at all is the way and the nature of its activities. Parts of a car's engine have to engage in specific activities in a specific type of organization if the engine is to work. Some of these parts can be replaced with different parts. If these parts engage in the same activities, it is still the same kind of mechanism. Consider the replacement of damaged tissue with healthy, although different, tissue. For example, portions of torn ligaments can be replaced with hamstring or some other tissue. If an entity is replaced with another entity in a mechanism, but it engages in the same type of activity, the mechanism continues to operate as it did. Activities are always activities of some entities, but certain kinds of activities need not be metaphysically bound to particular kinds of entities. A heart can continue to operate normally even if it has one or more bypasses made of a synthetic material.

[42] See Bogen (2008) for a discussion on conditions under which various productive activities fall under the concept of activity.

Glennan in his (1996) states that interactions between entities are determined by direct causal laws. In a later paper from 2002, he disregarded the reference to causal laws and accepted a view similar to Woodward's interventionist account of causation. I have presented the core ideas of Woodward's interventionist account in section 1.5. Recall how the central idea of Woodward's account is that causal relations are relations between variables that can be exploited or manipulated by intervention. For Woodward, two variables are in a causal relation if some experimental intervention on the value of a variable taken as a cause would generate a change in the value of a variable taken as an effect. An intervention is first and foremost a counterfactual notion which includes idealized experimental manipulations. Thus, two variables stand in a cause-effect relation when there is a covariant relation between them which is, most importantly, exploitable for experimental manipulation (at least, in principle).

Consider now the example of the lac operon model for *E. coli* discussed by Woodward in his (2002). If lactose is not present in the environment, a regulatory gene produces a repressor protein which binds to the operators and prevents transcription of lactose metabolizing enzymes. When lactose is present, an isomer of lactose, allolactose, is produced and binds to the repressor protein, preventing it from binding to the operator. Woodward rightfully takes this as a case of double prevention and argues that we cannot establish the productive relationship between allolactose and enzyme production by appealing to MDC's activity approach:

> And while we can perhaps use MDC's list of bottom out activities to describe the productive relationships between individual steps in the above process, it is far less obvious how to use this list to capture the idea that there is an overall productive relationship between allolactose and enzyme production without explicitly invoking the idea of counterfactual dependence. To begin with, the overall relationship between allolactose and enzyme production does not seem to fall into any of the categories on MDC's list.
>
> Woodward 2002: 372

An interventionist account of causation, according to Woodward, is enough to explain the functioning of a mechanism's parts, together with spatiotemporal information and the fine-tunedness of organization (Woodward 2013). Although not referring to possible words in order to establish truth conditions for counterfactuals, interventionism still defines causal relation as

if the intervention is performed – that is, it is still a counterfactual theory of causation.[43] Therefore, Illari and Williamson argue that if mechanisms are to be "real" and "local", the metaphysical account of causal relations in mechanisms has to acknowledge such relations as actual and not counterfactual (in the sense that they are dependent on possible worlds or on any other interpretation of counterfactual dependence). Mechanisms, they argue, require "active metaphysics such as Cartwright's capacities approach, a powers approach, or an activities approach" (Illari and Williamson 2011: 841). Whether or not the activity approach can be plausibly held from the metaphysical standpoint is a matter of debate. However, it does have at least one advantage over monist ontological accounts based on some difference-making theory of causation: parts of a mechanism can perform activities which do not affect other parts; for example, an entity *travelling* through a signaling channel or an entity rotating on its axis.

Is there a way to account for both counterfactual causal relations and productive, active causal relations within one account of mechanisms? In his (2004), Tabery offers an account of mechanisms which includes both interactions and activities. A similar approach can be found in Glennan's more recent writings too. In (2009) and in (2017) Glennan claims that causal relations in any mechanism come in two varieties: productive relations and relations of causal relevance. Similar to Glennan's argument, I have claimed in my (2021) that to retain all three aspects of mechanisms (entities, activities/interactions, organization) in an account of ontic mechanism, both types of causal relation within a mechanism – some sort of difference-making relation which corresponds to the causal role of the organization within a mechanism (for example, a counterfactual theory or interventionism) and some sort of productive relation between the mechanism's parts – are necessary. However, this presents a problem for the ontic conception of explanation and the OM thesis. I will present my arguments for this claim in the section 2.6.

2.3.3. Organization

Why and how does the inner organization of mechanisms matter? It is possible for the same entities to engage in different activities. It is possible that different entities engage in the same activities. It is even possible for the same entities to engage in the same activities yet in

[43] It does not matter if an intervention is not practically possible. It is enough for an intervention to be possible in principle. A variable, whatever it stands for, *ought to be manipulable*.

a different temporal order and which, then, results in a different effect. Examples of these organizational features can be very simple. Consider one provided by Glennan and Illari: "the very same resistors or capacitors can exhibit very different resistance or capacitance depending upon whether they are wired in parallel or in series" (Glennan and Illari 2018: 95). Another reason why we should consider organization as a constitutive aspect of any mechanism is that organizational aspects of mechanisms are not conditional upon specific entities and specific activities of the mechanism. Organizational features of mechanisms can be shared by a number of different mechanisms from different special sciences (or different ontological domain if you like) with completely different entities and activities.

One simple way to define mechanisms by referring to their organization is by contrasting mechanisms to aggregative structures. Wimsatt's four conditions for aggregative structures are probably the best known so I will use them here to stipulate how organization set mechanisms apart from mere aggregations. Wimsatt's account of aggregative structures is motivated by discussions on emergent properties and structures, and the connected issue of reductionism of special sciences. Emergent properties are pervasive in the sciences, according to Wimsatt. However, providing a definition or necessary and sufficient conditions for emergent properties is not straightforward. They are products of "organizational interdependence", but this comes in a variety of possible forms (Wimsatt 1997: S375). Therefore, he provides criteria to identify cases where emergence clearly fails, that is, criteria of aggregative structures. First, parts of an aggregative structure can be rearranged in any way without change in the overall output of the system or its constitution. Second, aggregative systems differ only in qualitative or quantitative addition and subtraction of their parts. Third, aggregative systems can be decomposed and rearranged and still continue to exhibit invariance in outputs or constitution. Finally, parts in aggregative systems do not enter in cooperative or inhibitory interactions. That is, there are no consequences on the overall output of the system which is due to interactions between parts.

Obviously, mechanisms as they have been defined so far do not exhibit these four conditions. In other words, mechanisms exhibit certain interrelated interactions between their parts and their activities which prevent their characterization as aggregative structures. The absence of organizational relations and properties means that mechanisms are also absent. However, some mechanisms exhibit none of Wimsatt's for criteria of aggregative structures, while others might still exhibit some of them. The upshot is that although organization is a

defining aspect of mechanisms it is not a two-valued variable. Organization comes in degrees. Simply, some mechanisms are more organized than others. Some mechanisms possess a highly complex internal organization (e.g., mechanisms for gene expression) while others have simple organizational features (e.g., electric circuit mechanisms).

The organization of mechanisms can also be distinguished into two general types: horizontal and vertical organization (Glennan 2017). Horizontal organization refers to the relations between parts. Glennan calls this the *topological* organization of a mechanism: the "abstract pattern of connection between its parts" (Glennan 2017: 121). Similarly, Levy and Bechtel claim that such organizational features imply that: "(a) different components of the system make different contributions to the behavior and (b) the components' differential contributions are integrated, exhibiting specific interdependencies (i.e., each component interacts in particular ways with a particular subset of other components)" (Levy and Bechtel 2013: 244). Similar views are expressed by Craver where he calls this the "active organization": "Mechanisms, in contrast, are not mere static or spatial patterns of relations, but rather patterns of allowance, generation, prevention, production, and stimulation" (Craver 2007a: 136). Woodward's "fine-tunedness of organization" expresses the same idea: the synchronized work of a mechanism's parts produces the effect. Changing the causal contribution of a part changes the contribution of the whole. Changes in the patterns of causal connectivity change the overall output of a mechanism.

The second type of organization – vertical or constitutive organization – refers to entities and activities being ordered spatially and temporally. Mechanisms can have numerous parts or there can only be a few of them. Sometimes, a part's spatial location and orientation is crucial and sometimes it is not. Some parts, on the other hand, are synthesized during the course of a mechanism's continued work. The timing of their synthesis or the timing of their entering into causal relations is vital for the mechanism's performance. Recall from section 2.2. that MDC characterize mechanisms in terms of start to end conditions. Each of the steps in the production of a phenomenon requires at least some kind of temporal order. Therefore, the properties of parts (causal, spatial, and temporal) give rise to and "sustain" their active organization (Craver 2007a: 137). If mechanisms are mechanisms for some kind of a behavior, then the regular production of an effect by a mechanism is a distinct feature of the mechanism. The notion of regularity here needs a little more interpretation since one-off mechanisms are ubiquitous in biology and elsewhere. A bomb is a mechanism too. However, such a mechanism,

if it works properly, produces its effect only once: an explosion. What makes it capable of regular production of the explosion is that such an organization of parts and their interactions can and often will cause an explosion. Nevertheless, biological mechanisms (or social, cultural, or economical ones, for that matter) are usually organized so as to continuously produce their effects. Protein synthesis mechanisms produce proteins continuously when the conditions are satisfied.

There are two other organizational features I will briefly mention here: hierarchy and modularity. As already stated, mechanistic philosophy is both a worldview and an account of scientific practice. In that regard, the mechanistic worldview is a view of hierarchies and layers. Similarly, mechanistic explanation and its strategies and methodologies are *decompositional*, meaning that they seek to explain the phenomenon by looking downwards and decomposing the phenomenon into its causal or constitutive parts. However, a mechanism's parts are often themselves lower-level mechanisms and can be decomposed and described mechanistically. The hierarchy of mechanisms is both a metaphysical and epistemological issue. Levels are "construed as both ontic levels of mechanistic organization and as epistemic levels of analysis" (Wright and Bechtel 2006:55). An important caveat in thinking about constitutive mechanisms and inter-level relations, however, is in the methodology of discovering the constitutive relevance and interlevel relations as these have been developed along the lines of the interventionist account of causality. But, as it has been claimed in various places, testing causal relations cannot explain or provide reasons to assume constitutive relations. The most known attempts come from Craver and his mutual manipulability approach and Craver and Bechtel's mechanistically mediated effects. I will not discuss this problem here since it is not a pressing matter for the moment. For discussion, see, for example, Craver (2007a, 2007b), Craver & Bechtel (2007), Couch 2011, Leuridan (2012), Povich and Craver (2017), Kästner and Andersen (2018), and Craver, Povich and Glennan (2021).

The modularity of mechanisms is another key aspect of mechanistic philosophy. As with levels, modularity can also be taken as a metaphysical thesis concerning the organizational constitution of mechanisms or as an epistemologically necessary precondition for mechanistic explanation. Mechanisms are thought to be composed of modular assemblies. It presumes that we can alter or stop the causal activities or outputs of a single part of a mechanism without affecting the individual causal contributions of its other parts. The interventionist account of causation, which is supposed to reflect the scientific rationale and methodology for discovering

the causal contributions of specific variables, presumes modularity. It cannot work if a system of variables is not modular. As Woodward states, "The basic idea that I want to defend is that the components of a mechanism *should be* independent in the sense that it *should be* possible in principle to intervene to change or interfere with the behavior of one component without necessarily interfering with the behavior of others" (Woodward 2002: S374, emphasis added). Simon presents an interesting argument to argue that natural selection favors (or at least ought to favor) modular assemblies.[44] Such an organization is more resilient to malfunctions and perturbations: parts are then easily replaceable. Nonetheless, whether mechanisms are modular or not is a matter of empirical investigation. It cannot be assumed *a priori* in our definition of ontological mechanisms since it may come out that only a few or perhaps none of the mechanisms found in biological and medical sciences are indeed completely modular.

2.4. Epistemic Mechanicism

If OM is a thesis about things in the world, EM is a thesis about how scientists try to represent that world. It asserts that explanations in the life sciences are mainly mechanistic explanations. The thesis does not argue for a uniformity of scientific explanations across sciences or even within a single scientific field. Mechanistic philosophers acknowledge that there are different kinds of scientific explanation, and it is quite possible that some of them are not causal.[45] However, the EM thesis does argue that causal explanations from different sciences and especially in the life sciences are of a mechanistic type. The thesis has both descriptive and normative features. Applying the bottom-up methodology, its descriptive claim states that mechanistic explanation provides a description of a causal structure, deconstructed into its component parts, their activities, and interactions, and the organization of these into the overall mechanism. Such a description explains how these structures are causally or constitutively responsible for phenomena. The normative part of the EM thesis provides the necessary features of such an explanation, and a set of criteria to determine its quality.

As mentioned, mechanisms are distinguished as being either *constitutive* vs. *etiological* (Craver 2007a, Glennan 2009) or in different terminology, *vertical* vs. *horizontal* (Kincaid

[44] See Simon's example with two watchmakers Hora and Tempus from his (1962).

[45] For non-causal explanations in science consider, among others, Lange (2016) and Reutlinger and Saatsi (ed.) (2018).

2011).[46] We can think of constitutive or vertical mechanisms as structures that are simply constitutive of a given object or phenomenon. For example, to explain what a heart is and how it works (explanandum) means to provide an explanation of its constitutive mechanism (explanans). The phenomenon (e.g., heart) just is the mechanism that needs to be explained. Furthermore, it is often emphasized that a constitutive mechanism's parts are often themselves lower-level mechanisms. In that sense, mechanistic philosophy takes phenomena as being hierarchically decomposable into lower-level mechanisms. A description of constitutive mechanism, then, is something like this:

> The heart is itself part of a larger mechanism, the circulatory system, that includes such parts as veins, arteries, and the blood itself. Parts differ in their roles with respect to particular operations; for example, the chambers of the heart play an active role in the operations of contracting and relaxing whereas the blood plays a passive role in the same operations (it undergoes change of location). The various components must be both spatially and temporally organized such that blood can flow on each side from atrium to valve to ventricle to valve to aorta or pulmonary artery into the rest of the circulatory system, as suggested by the arrows. At least as important, the operations must be precisely timed to achieve an orchestrated effect.
>
> Bechtel and Abrahamsen (2005: 424)

Horizontal mechanisms, on the other hand, lead to, produce, or bring about the effects. Horizontal or etiological mechanisms, then, can also be understood as constituting Hall's productive causes: to be a horizontal mechanism amounts to having the right kind of internal structure which is identified as a union of minimally sufficient sets for *e* in every time between *t* and *t'*. For example, heart failure is a condition where the heart, due to different etiologies (and where hypertension is just one of them), simply cannot pump blood sufficiently well. Therefore, a mechanistic explanation of heart failure should provide descriptions of entities and their interactions at every step of the causal chain connecting, for example, hypertension and heart failure. An explanation employing a horizontal mechanism of heart failure due to essential hypertension then could be something like this:

[46] The term "etiological" here, however, does not correspond to the term "etiology" from the medical literature, discussed in the first chapter.

> Typically, a rise in blood pressure is sensed by the smooth muscles' lining vessel, which releases nitric oxide leading to vessel dilation and reduced resistance. Elasticity in blood vessels is gradually lost due to aging making it increasingly difficult to adjust vessel diameter and account for changes in blood pressure. These vessels are constantly constricted causing increased pressure that erodes the cardiomyocytes and increases fibroblasts, collagen, and hypertrophy of cardiac tissue.
>
> Capote et al 2015: 38

Whether we are talking about constitutive/vertical or etiological/horizontal mechanisms, *Epistemic mechanicism* is all about modeling. Hence, the main claim of the EM thesis is the following:

EM1: All mechanistic explanations are models of mechanisms.

Mechanistic explanations are representations of mechanisms, and representations of mechanisms are models of mechanisms. What, then, is a model of a mechanism and what does the construction of a model of mechanism amount to? Before answering this question, the notions of model and model construction need clarification since there has been quite a rich discussion about model construction and models in the philosophy of science for the past fifty or so years. To avoid certain ambiguities concerning the term model, it is better to resolve possible misunderstandings or potential problems right at the beginning. I will not, however, engage in the discussion about the relation between models and theories or scientific laws or about models and similarity/informativeness since these would lead me far astray.

2.4.1. Models

Models are present everywhere in science. Consider Watson and Crick's model of DNA or the Ptolemaic model of the solar system. In biology especially, one rarely hears of theories. Models on the other hand, a go-to explanatory term in biology (and especially in molecular biology). Giere claims that models are not something "ancillary" in science but rather occupy the central role in scientific accounts of the world (Giere 1999). Usually, models are taken to be representations. This is a view on models I will accept here too. Two positions on models are then predominantly argued for. Models are either similar to their targets or they convey some information about the part or parts of the world they are representing or referring to

(without being similar). For example, Glennan writes: "Models are only models when they are used to represent something, and it is the modeler's act of interpreting the model— asserting the similarity between some aspect of the model and some aspect of the target—that determines what kind of model it is" (Glennan 2017: 67, 68). Similarly, Giere (2004) claims that models are abstract objects specifically intended to be representations: "What is special about models is that they are designed so that elements of the model can be identified with features of the real world" (Giere 2004: 747). They are used as "tools for *representing the world*" (Giere 1999: 44). Bolinska (2013) takes scientific models to be a subclass of a more general notion – epistemic representations. Something is an epistemic representation "of a given target system if and only if it is a tool for gaining information about this system" (Bolinska 2013: 221). The informativeness of epistemic representations about their target systems is a feature that distinguishes scientific models as epistemic representations rather than some other kinds of representations (e.g., a painting or a sculpture) although it could turn out not to be the only one applicable to the system in question. Nonetheless, their informativeness is their specific function or use.[47] Let me add an additional claim here:

> **M*:** Models are representations which have the capacity to convey some information about some portion, part, or aspect of the world – the model's target systems.

Models are not abstract objects. Rather, they represent something in an abstracted way. They are abstract representations in the sense that they intentionally leave out a lot of information about the phenomenon. Consider how ball-on-stick models in chemistry are supposed to represent atoms and bonds between them. Atoms are represented as spheres while bonds are represented as rods. The properties of the spheres (such as their shape and color), the ratio between the diameters of atoms and the length of the rods are obviously used to give a simplified representation of molecular structure – in a nutshell, it is a false representation of the phenomenon. But these features of ball-on-stick models have their specific purpose. The features of any model are always there for a specific purpose whether the purpose is explanatory or something else. Giere especially notes how the intentions and purposes of a modeler play a crucial role in model construction. He develops his account of models based on

[47] There is a rich discussion on scientific models and representation in the philosophical literature for which I do not have space to engage with here. See Magnani and Bertolotti (eds.) *Springer Handbook of Model-Based Science* (2017). For analysis and discussion on informativeness of models see Suarez (2004).

the following proposition: "S uses X to represent W for purposes P" (Giere 2004: 743). Here, S represents any group, community or an individual.

The same phenomenon can be represented by numerous and vastly different models. Often, the particular characteristics and features of a model will depend on the goals of a particular scientific or educational (or any other) group or community. The purposes of a model are various: educational, explanatory, explorative, predictive etc. Similar to Giere, Teller also emphasizes the purpose of the model intended by the modeler: "in principle, anything can be a model, and that what makes a thing a model is the fact that it is regarded or used as a representation of something by the model users" (2001: 397). Models are developed and refined gradually as our knowledge of the phenomenon grows but, in the end, how the model looks is characteristically defined by its intended purpose. Although false, ball-on-stick models are still in use precisely because they offer an easy way to grasp the characteristics of molecules that we usually take to be important.

How does a model represent a portion of the real world if it is an intentional misrepresentation? Considering some general views on the relation between models and the real world, Giere's analogy of models with maps can be helpful here (Giere 1999). Maps represent some area of space. They involve a lot of details or features of that area but they also leave out a considerable amount. For Giere, then, a map represents that portion of the real world by being spatially similar to it (among other similarities between features of the real world and a map). Consider a map of any city. You will see streets, roads, avenues, parks, bridges and possibly some buildings on a 2D plane. It will help you to get an understanding of the spatial layout of the city and to estimate distances between different neighborhoods or between different streets. It can definitely help you to get from point A to point B when you find yourself in the actual city. But a map will certainly not be a detailed representation of neighborhood, city, or area. Far from it. No map looks exactly the same as the geographical location it represents, and similarly, no model looks exactly like the real thing in the world it represents. There is no perfect model.[48] Every model is a mixture of abstractions and idealizations.

[48] "The only PERFECT model of the word, perfect in every little detail, is, of course, the world itself" (Teller 2001: 410)

Nonetheless, every model is in some respect similar to the thing it represents. How could it be explanatory or informative otherwise?[49]

What can be a model? Simply, anything can be a model as long as it represents some portion of the real world, that is, if it fulfills its purpose (as intended by the modeler) of representation in some specific way. Therefore, models can be diverse. Frigg and Hartmann (2020) distinguish between scale models, analogical models, idealized models, toy models, minimal models, phenomenological models, exploratory models, and models of data. Giere (2004) distinguishes between physical models, scale models, analogue models, and mathematical models but recognizes that the list could be much larger. Nonetheless, what all models have in common is that they are intended as representations of some portion of the real world.

2.4.2. Models of mechanisms

Recall that the distinguishing feature of the deductive-nomological model of the CL account of scientific explanation is that it takes the form of a deductive argument where the premises are the explanans (initial conditions and laws of nature) and the conclusion is the explanandum (a phenomenon). On the other hand, EM states that mechanistic explanations do not have some universal form in the way that DN explanations have (i.e., premises of a deductive argument). Although every mechanistic explanation is a model, a model of a mechanism might take numerous forms. For example, a mechanistic explanation can come in visual, textual, diagrammatic, or mathematical form.

Consider a mechanistic representation of the human heart. We could give a textual explanation of how the heart works, including its relevant parts, such as right and left atria, right and left ventricles, pulmonary and aortic valves, pulmonary artery, and pulmonary vein, and the activities of these parts. But we could also make a video representing how those parts of the heart work and contribute to its overall working. The video, for example, could represent

[49] Going into a discussion of similarity would be too philosophically demanding, time consuming, and exhausting. I will only mention that instead of resolving the issue of what similarity is in general, Giere rather focuses on specifying the respects in which a model is similar to the real thing and the degree of similarity it has in those respects. See Giere (1999), Suarez (2003), and Poznic (2016) for the more detailed discussions on this matter.

how right and left atria receive the blood that comes into the heart, and how right and left ventricles pump the blood out of the heart, and so on. It might represent the heart in three dimensions and therefore show the spatial relations between its parts. It could also represent temporal relations or the temporal order of the activities and interactions of these parts. Such a representation of the heart could be detailed and very informative. But the heart can also be represented pictorially, as is usual in medical textbooks (as well as other educational materials). A pictorial representation will include some component parts and might indicate some of the activities and interactions of these parts. These causal relations could be represented, for example, with arrows, denoting the directions of activities and interactions (for example, the direction of the blood flow). Such a representation of the heart will sometimes carry different kinds of information or additional information that, for example, textual representation does not contain. An example is shown in Figure 7. The model represents some of the component parts, and while it leaves out most of the interactions and activities of those parts, it indicates the direction of flow by using arrows. In sum, it offers some insight into the spatial organization of parts and at least this one type of activity.

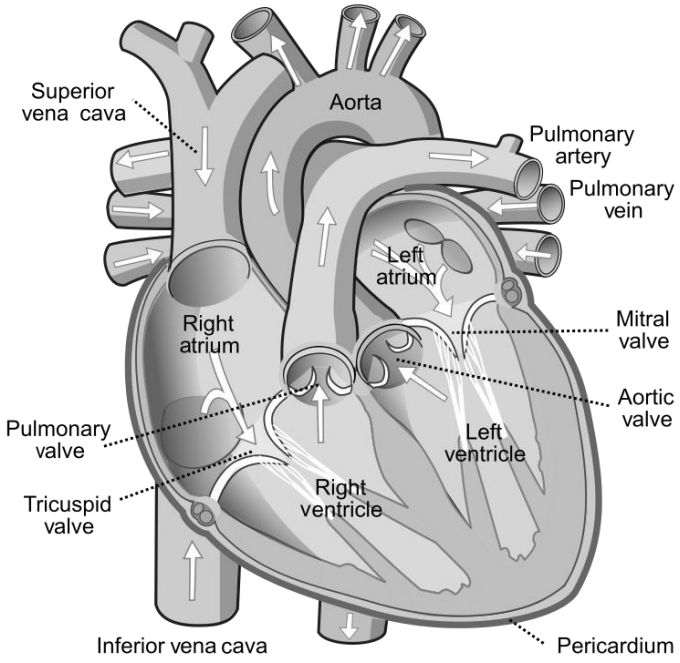


Figure 7. The usual textbook model of the heart.[50]

[50] By Wapcaplet – Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=830253

On the other hand, a network or mapping representation of some genetic mechanism will carry some information that diagrammatic and pictorial representations cannot convey. Mechanisms can also be described by sets of equations if these equations and their variables correspond to the parts and activities of the mechanism. Models of mechanisms might also take the form of causal Bayesian nets, representing causal organization within the mechanisms. Such models leave out certain features of the mechanism such as spatial and temporal relations, as well as numerous other features of its parts and organization, but they give us a representation of causal organization that, perhaps, textual, pictorial, and other representations of the mechanism cannot express.

Not all models are explanatory. A model of data does not explain any real-world phenomenon. But, according to some mechanists, there are some models that of a phenomenon and that carry some explanatory force, yet without being mechanistic explanations or mechanistic models. Many mechanistic philosophers have pointed to the difference between phenomenal models and mechanistic models in order to emphasize necessary features of mechanistic models. Phenomenal models do not provide any causal explanation, they are just descriptions (or redescriptions) of a phenomenon (e.g., as in Craver 2006, 2007a, Kaplan 2011, Kaplan and Craver 2011). Craver has probably written the most on this distinction and has discussed it in quite a few of his publications. In his (2006), he writes: "A model can be richly phenomenally adequate and non-explanatory. This is the take-home lesson of the several decades of attack on covering-law models of explanation at the hands of advocates of causal–mechanical models of explanation: merely subsuming a phenomenon under a set of generalizations or an abstract model does not suffice to explain it" (Craver 2006: 357, 358). Models that do not explain yet say something about a phenomenon are characterized by Kaplan and Craver as models that "save the phenomena" rather than explain the phenomena (Kaplan and Craver 2011). In a way, phenomenal models describe or redescribe what the phenomenon is, what it consists of, or how the mechanism behaves under different circumstances. Phenomenal models are Cummins-type functional explanations (Cummins 1975). They decompose a system's behavior into the functions and capacities needed for that overall behavior, but they are not decomposed structurally in the sense in which mechanistic explanations are.

In his (2006) Craver discuses Hodgkin and Huxley's model of action potentials as an example of a phenomenal model (Figure 8).[51] The model is composed of several equations where "the central one" provides a description of how the total current crossing the membrane changes according to changes in capacitive current, potassium current, sodium current, and the sum of smaller currents of other ions (leakage current). Craver supports his argument for the lack of explanatory force in the model in part on Hodgkin and Huxley's own admission that they do not intend the model to be a causal explanation in which elements of the model correspond to something physical in the real mechanism. That is, Hodgkin and Huxley claim that they have no doubt that an equally satisfactory model for predicting these changes could have been achieved with a different equation or set of equations. They argue that the equation they have come up with summarizes observations from experiments conducted over a few decades.

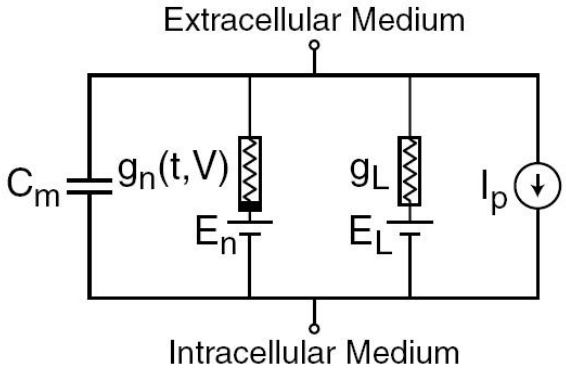


Figure 8. The Hodgkin and Huxley model of action potential.[52] The equation for total current is $I = C_M dV/dt + G_K n^4(V - VK) + G_{Na} m^3 h(V - V_{Na}) + G_l(V - V_l)$

Craver, echoing Hodgkin and Huxley, writes: "In the HH model, commitments about underlying mechanisms are replaced by mathematical constructs that save the phenomena […] of the action potential much like Ptolemy's epicycles and deferents save the apparent motion of the planets through the night sky. The equations, in short, do not show *how* the membrane

[51] Alan Hodgkin and Andrew Huxley received the Nobel Prize in Physiology or Medicine in 1963 for their work on action potentials.

[52] CC BY-SA 2.5, https://commons.wikimedia.org/w/index.php?curid=642185

changes its permeability" (Craver 2006: 364). Again, mechanistic explanation ought to describe how parts and their interaction, in virtue of their properties and organization, bring about the phenomenon. Here, however, instead of a "how" model, we have a "that" model. The model predicts the behavior of the mechanism but does not explain it.

Nevertheless, in his (2007a), and later in a paper coauthored with Kaplan in 2011, Craver argues that the HH model in fact is not entirely a phenomenal model. Taken by itself, the equation does not provide a "how" explanation. But, supplemented with additional information about the correspondence of variables to the mechanism's parts, the model becomes partially explanatory. Craver claims that this is, in fact, what Hodgkin and Huxley had at their disposal when they were thinking about the model: "They knew, for example, that the action potential is produced by changes in membrane permeability, and they knew that ions flux across the membrane toward their equilibrium potentials; they also knew that this flux of ions constitutes a transmembrane current. This background sketch of a mechanism does provide a partial explanation (an explanation sketch) for how neurons generate action potentials because it reveals some of the components of the mechanism, some of their properties, and some of their activities" (Craver 2007a: 56). The take home from this discussion is that the extent to which a model of mechanism is explanatory is a question of degree. One model of a mechanism may convey more information about a phenomenon than some other model. As our knowledge about parts, activities, interactions, and organization grows so should grow the explanatory and predictive powers of our models. Although Hodgkin and Huxley had a successful predictive model, according to Craver, they also had a sketch of the mechanism of action potentials – an incomplete mechanistic explanation, but still a mechanistic explanation. As Glennan remarks in this connection, "The point here is that sketchy mechanistic models are different from phenomenal models, and sketchy models explain, albeit sketchily" (Glennan 2017: 67). However, Kaplan and Craver conclude that models are mechanistic only when they describe the real structure bringing about a phenomenon. Kaplan and Craver call this requirement "model-to-mechanism-mapping" or "3M" (Kaplan and Craver 2011: 602). In a way, this is similar to what Glennan has in mind when he claims that every element in a mechanistic explanation has to correspond to an element from the real mechanism. That is, a part cannot exist only as a place holder in a model of a mechanism. It has to be a real, robust thing.

The idea that the explanatoriness of models of mechanisms come in degrees is present in the works of many mechanistic philosophers. For example, concerning the distinction between a mechanistic or some non-mechanistic models (like the above discussed phenomenal models), Woodward discusses "whether we should think of the mechanical/non-mechanical contrast as a sharp dichotomy, or instead more in the nature of a graded (perhaps multi-dimensional) continuum according to which some explanations are more 'mechanical' than others (in various respects)" (Woodward 2013: 41). Therefore, it might be better to think in terms of a continuum, where some mechanistic explanations are more mechanistic than others. This shift from less to more mechanistic is perhaps best discussed in MDC (2000) and in Craver and Darden (2013) where these authors represent this gradual shift as the move from (i) mechanism sketch to mechanism schema, and from (ii) how-possibly to how-plausibly and how-accurately mechanistic explanations. (i) refers to the amount of detail in a representation of mechanism, while (ii) is perhaps best expressed as the degree of evidential support.

A how-possibly model is a proposition about a possible kind of mechanism underlying a phenomenon. At this stage, a certain rough draft is proposed where one is not concerned whether parts and their operations and interactions in a model represent the real mechanism in question. How-possibly models of mechanisms are, according to Craver, "loosely constrained conjectures" (Craver 2006: 361). Such models serve only as an initial guide for further investigations. They provide "invaluable heuristic information", and, as such, serve their purpose to narrow the spectrum of possible mechanisms (Kaplan 2011: 353). A how-possibly model becomes a how-plausibly and eventually a how-actually model when empirical evidence confirms, not only that such a model can bring about the explanandum but that it indeed does so (since there might be several different models of mechanisms that are equally capable of producing the phenomenon). A mechanism schema, on the other hand, is a more or less complete representation of a mechanism. It includes entities, activities, and organizational features with enough details so that there are no "placeholders" terms. A mechanism sketch, although still a representation of a mechanism is missing important details; there are gaps in our understanding of the mechanism's features. The gaps are filled with placeholder terms that designate an unknown factor (whether a part or an activity).

For illustration, recall the story from the previous chapter about how NO was identified as EDRF. It was known that vasodilation was the effect of smooth muscle cell relaxation which, in turn, was the effect of a specific factor released in the endothelium. The model of smooth

muscle cell relaxation was incomplete since this crucial factor was unknown – it was a sketch rather than a schema. The model included a black box. Nevertheless, it was suspected that it was a gaseous compound acting as a second messenger. Since it was observed that it is generated in the endothelium it was called endothelium-derived relaxant factor – EDRF. Although scientists had a mechanistic explanation of smooth muscle cell relaxation, the explanation was not complete. It included filler terms, a placeholder – EDRF. However, the model itself was a how-actually model since it was confirmed that not only could the model account for vasodilation, but indeed it gave the very mechanism underlying vasodilation. The model only lacked the identification of its crucial part – EDRF. Finally, the identification of nitric oxide (NO) as EDRF by the late 1980s was due to the combination and convergence of research from three different areas – research on the immune system, cardiovascular system, and nervous system (Lancaster 2017). This example shows that as the research progressed, the black box in the mechanism of smooth muscle cell relaxation was replaced by NO. Its properties and its role in the causal pathway leading to an increase of cGMP and, consequently, vasodilation, made it possible to finally construct a supposedly complete mechanistic explanation of the phenomenon. Craver and Darden illustratively describe this as the process of turning a black box into a grey box, and finally to a glass box.

2.4.3. Abstraction and idealization

An indispensable feature of the discussion on models of mechanisms concerns the degree of abstraction and idealization in mechanistic explanations. Abstraction and idealization are methods that either simplify or distort a representation of a mechanism. The discussion in the mechanistic literature on this issue has one particular and one general concern. The particular issue is concerned with the use of abstraction and idealization in modeling mechanisms specifically, while the more general concern is what exactly assumes the role of explanans in scientific explanation (the ontic/epistemic dichotomy). Let us consider how these features are related to the explanatory success of a model.

Abstraction, of course, does not refer to abstract entities or objects. As already stated, both abstraction and idealization are properties or features of models of mechanisms, not real mechanisms themselves. They are methods, among others, that modelers use to represent mechanisms. Remember that all models are abstract in the sense that they are never perfect

representations of a phenomenon. They always leave out some information. Abstraction, then, is an intentional sparsity or omission of detail in a model of mechanism. A similar view on abstraction can also be found in Levy (2021). He defines abstraction as the level or degree of detail in a model. Models that have a high degree of abstraction are still true representations of a mechanism, albeit simplified.

Consider again a textbook model of the heart, such as a pictorial representation. Such a model does not lie about the thing it represents, so to say. A standard pictorial representation of the heart does not show anything that is not present in the "real thing". It does intentionally leave out a lot of features that hearts usually possess (parts, their properties, their interactions, and their organization), but such a model still leads to a representation of the heart that is supposed to be true. Idealization, on the other hand, intentionally misrepresents or distorts a phenomenon. Models that are idealized are not true. Idealized parts or idealized causal relations between parts do not exist in the real world. Consider certain explanations from ecology, population genetics, or evolutionary biology. These will frequently include infinitely large populations of swallows, wolves, or rabbits. A model of molecular mechanism may distort the causal relations between proteins and operators. These features of models help us to either move away from the details or to simplify the details for the purposes of understanding, or as Levy and Bechtel claim, they "expedite analysis and understanding" (Levy and Bechtel 2013: 243).

Many philosophers argue that abstraction and idealization are characterized by the distinction between truth and falsehood (e.g., in Jones 2005, Godfrey-Smith 2009, or Levy and Bechtel 2013). Abstract models are conceived as still being true representations of the portion of the world they represent, while idealized models are necessarily false. Jones, for example, takes idealization as "assertion of falsehood" while abstraction is "omission of a truth" (Jones 2005: 175). Similar views can be found in, for example, Godfrey-Smith (2009) and Levy and Bechtel (2013). Portides (2021), however, has a different and interesting view that breaks away from the typical distinction between truth and falsehood. He ties both abstraction and idealization to the selective attention of a modeler to certain features of a mechanism. The two aspects are then linked to cognitive processes or the intentions of a modeler rather than anything conceptual, logical, epistemic, or anything else unrelated to the purpose of a model. This echoes Giere's conception of a model where a model is always a representation of something for a certain purpose. Some features of a mechanism are intentionally abstracted and/or idealized

since this serves some purpose. The purpose sets the thresholds of abstraction and idealization above which the model stops being similar to or informative about the target system.

No matter how we are to understand abstraction and idealization, the question of a relation between the completeness of a model of a mechanism and its explanatory force remains. If mechanistic explanations describe how phenomena are produced or how they are constituted by mechanisms, then it seems that abstraction and idealization are troublesome features for the characterization of a good, and especially a complete, mechanistic explanation. Hence, the following questions should be addressed. Do we always strive for more detail when trying to understand a mechanism? Does an explanation of a mechanism always carry more explanatory force when we have more details about parts, causal relations, and the organization of a mechanism? When is a mechanistic explanation complete and when is a mechanistic explanation good?

For now, let me continue with the description of the three theses of "The New Mechanistic Philosophy". After presenting the third thesis in the following section, I will give my account on the relation between ontological mechanisms, models of mechanisms, and mechanistic methodology. I will also present my view on the criteria of a good mechanistic explanation.

2.5. Methodological Mechanicism

The third thesis of "The New Mechanistic Philosophy" – *Methodological mechanicism* – is a set of claims from the mechanistic literature expressing arguments for a distinct mechanistic way of doing science (and life sciences in particular). That is, it is a set of claims expressing a specific strategic approach and methodology labeled "mechanistic", which includes necessary or at least distinctive criteria for constructing mechanistic models. Most if not all mechanistic philosophers accept (to varying degrees) all three theses and, as I have been claiming, they are often considered interchangeably in the mechanistic literature: the end product of doing science mechanistically ought to be a mechanistic explanation of a phenomenon and mechanistic explanation ought to be a representation of a mechanism as it is found in nature. But, still, each of these theses represents just one aspect of the discussion and the use of the term mechanism.

2.5.1. Mechanistic strategy of inquiry

There are only a few works in the mechanistic literature which have been explicitly or mostly concentrated on the mechanistic methodology in science. Most of the discussion on the matter, however, is present in Bechtel and Richardson (1993), Darden and Craver (2002), Darden (2006), and Craver and Darden (2013). For example, in (2002) Darden and Craver write: "Focusing centrally on mechanisms provides new ways of thinking about discovery, interfield integration, and reasoning strategies for scientific change" (Darden and Craver 2002: 2). Darden's and Bechtel and Richardson's books deal specifically with these strategies used in mechanistic scientific methodology.

Similar to the views of Darden and Craver, Ioannidis and Psillos (2018) claim that being committed to mechanisms means taking a certain methodological stance. But in doing so, Ioannidis and Psillos's argument resembles Moss's view that the term "mechanism", as it is used in the life sciences, does not have a definite meaning. Rather, mechanism implies a methodology above anything else. Most likely, the majority of scientists from the life sciences would agree with one or maybe all of the definitions of mechanism from the philosophical literature, but it is an unfounded philosophers' assumption that they in fact have any of those definitions in their minds when actually doing science and constructing explanations within their respective domains.

Ioannidis and Psillos presume that the term mechanism is nothing more than "a certain theory-described causal pathway" (Ioannidis and Psillos 2018: 2). First, what is a causal pathway for Ioannidis and Psillos? For a start, it does not correspond to Ross's notion of causal pathway. In their view, a causal pathway is any process that can be characterized as "a regular sequence of events and difference-making relations among its constituents" (Ioannidis and Psillos 2018: 2). Second, what does "theory-described" mean? A certain causal pathway can be approached from different perspectives, or different fields of science, and an account of a phenomenon can be given by using theoretical terms from those various field. Consider the example they discuss – the process of apoptosis or programmed cell death. As they argue, scientists have given us accounts of apoptosis from different perspectives, conditional on the motivations for describing and understanding the phenomenon, varying from the cytological to the biochemical point of view. The mechanistic stance, then, refers to the causal investigation of a phenomenon where "the end result is a highly informative theoretical description that embeds the pathway within the known physiological and biochemical functions of the

organism" (Ioannidis and Psillos 2018: 4). Such a view was already stated in an earlier paper by those authors. There, they argue for a minimalist interpretation of a mechanism where a mechanism is, again, a methodological thesis which "allows that the sought-after identification of the causal pathway by which a specific result is produced, is fully captured in the language of a specific theory, using deeply theory-laden concepts" (Ioannidis and Psillos 2017: 605). According to them, the choice of an approach towards a phenomenon influences the choice of theoretical language used to describe the phenomenon. Phenomena, they argue, are always described in certain theoretical terms from different points of view.

Levy identifies the mechanistic stance with a certain set of cognitive methods used to approach "a particular set of phenomena" (Levy 2012: 105). In his view, the stance refers to "a framework for representing and reasoning about complex systems" (Levy 2012: 104, 105). But, as I claimed in the second section of this chapter, frameworks for representing and reasoning is exactly what the EM thesis stands for. My MM thesis is closer to what Ioannidis and Psillos have in mind. Similarly, I take it that the first out of three meanings of mechanism that Nicholson identifies in his (2012) is supposed to refer to the same thing – a description of a methodological approach towards discovery and explanation rather than a thesis about the explanation itself.[53] In my view, Ioannidis and Psillos's arguments reflect the thesis of *Methodological mechanicism* the best. They provide certain criteria for a mechanistic explanation, and they discuss what it means to capture the phenomenon in the language of a specific theory, but their argument should be interpreted as making the case that the mechanistic approach to science is, first and foremost, a methodological stance and not a metaphysical or epistemological one. However, in their discussion, the EM and MM theses are still intertwined in a manner that prevents them from disambiguating the two as separate theses.[54]

[53] Nicholson's *Mechanism* meaning of mechanism is a "philosophical thesis that conceives living organisms as machines that can be completely explained in terms of the structure and interactions of their component parts" (Nicholson 2012: 152).

[54] As a short sidenote here, consider Ioannidis and Psillos's arguments in light of the discussion from the previous chapter. Ioannidis and Psillos's view resembles the arguments from Fiorentino and Dammann (2015), De Vreese et al. (2010) and my own presented in the discussion on the mechanistic approach to disease causation from Chapter I. Concerning medicine, then, the thesis can also refer to a possible strategy for understanding and explaining disease causation and regular physiological processes. It corresponds to the methodologies of the basic medical sciences in their characterization of phenomena, strategies used to acquire or gather data, ways of interpreting data, and finally, ways to get evidence from data. Taking a mechanistic stance means explaining a phenomenon in terms of its causes

To conclude, a closer look at the discussions from the mechanistic literature in philosophy reveals a distinct thesis underlying the notion of mechanism separated from the discussion on mechanistic explanation; a thesis or stance which is predominantly about the approach to phenomena, not about a correct epistemic framework for representing them nor criteria for grading the quality and success of such an explanation.

Methodological mechanicism is best described by the stages of inquiry in the process of discovery of a phenomenon. In the mechanistic literature we find this process divided and analyzed into several stages. For example, Darden's schema summarizes many of the accounts from the literature. She distinguishes the mechanistic methodological approach "into at least four stages: characterizing the phenomenon, constructing a schema, evaluating the schema, and revising the schema" (Darden 2018: 258). Here, however, I will only present the first step – identification and characterization of a phenomenon – and how it influences subsequent stages of an investigation. This discussion will then be placed in the context of mechanisms in medicine. I will discuss the further stages in a bit more detail in the last chapter, as a part of a discussion on relation between mechanistic explanation and prediction.

2.5.2. Functions and phenomena

In his (1971), Kauffman argued that a system can be decomposed differently depending on our interests and purposes. There can be more than one set of "sufficient conditions for the adequate description of the behavior" (Kauffman 1971: 258). Although it is possible that there is some "ultimate decomposition", there does not need to be a single one such "that all other decompositions are deducible from it" (Kauffman 1971: 259). However, for the decomposition or a description to work, "descriptions of parts and processes of one decomposition need only be compatible with and not deducible from the descriptions of parts and processes of a different composition" (Kauffman 1971: 259). But what are the constraints for this carving of nature into different decompositions or mechanisms? Are there limits to it?

on an intra-individual level. That is, taking a mechanistic stance is to provide an explanation of a phenomenon on the level of biological, chemical, and physical causes. It says that certain methodologies used in laboratory sciences are the best way to investigate those underlying biological, chemical, and physical causes of human health and disease.

Again, most if not all mechanistic philosophers accept that some mechanisms underlie functions. There are mechanisms for some regular behavior or for some function. Can it be claimed then that the phenomenon itself just is the function of a mechanism? Craver and Darden do not identify the phenomenon with the function but rather with "the ability to perform a function" (Craver and Darden 2013: 5). A mechanism does not necessarily need to implement or fulfill its function all the time in order to be a mechanism for the function (consider how it is also usually claimed that dispositional properties need not be constantly manifested in order for an object to possess them). There is some initial weight to this position. For example, a car engine is still a mechanism for transformation of chemical energy into mechanical energy, even if it is parked and out of use for some time. Also, a function in the mechanistic literature might sometimes mean the function of a part of a mechanism, and sometimes the function of a mechanism as a whole. But still, the question remains. What is the function of a car? What is the function of the heart or kidney? What is the function of an NMDA receptor? What do we mean when we say a mechanism for a function?

I will not engage in thorough discussion on functions since it has been one of the most discussed notions in the philosophy of science and the resolution of the debate, as usually, does not seem close.[55] But, one way or another, the notion of function cannot be avoided. Without going deep into the discussion about functions let me just mention the three most popular accounts here. The s*elected effects* account states that the function of a biological trait (in this case a mechanism or a part of a mechanism) of an organism is whatever the effects it has that were selected for by the processes of natural selection. Next, the *fitness* account says a trait's function is determined by its contribution to the fitness of an organism. Boorse's account of health and disease can be taken as operating within the fitness account. The *causal role* account stipulates that the function of a trait is its contribution to the overall effect of a system which it is a part of. Rather than discussing which account of function from the literature is the most plausible, we should, I think, consider what kind of an account of function do mechanistic philosophers have in mind and what kind of an account of function do medical scientists and practitioners have in mind in their practice.

Before I continue, I should address a possible issue. At this point, a reader may wonder why I include a discussion on functions and mechanisms in my methodological thesis, and not

[55] Consider Garson (2016) for the latest overview of discussion and positions on functions in philosophy.

in the ontological or epistemological thesis? The reason I consider this as a part of mechanistic methodology is because, ultimately, there are multiple ways to consider what a mechanism is and how to represent it, depending on the theory of function we accept. Garson makes a similar observation: "people might think about mechanisms slightly differently depending on how they think about functions" (Garson 2018: 108). Similarly, a rather different approach to discovery, explanation, and understanding can follow if a scientist considers that some biological phenomenon is a byproduct of a functional mechanism, or it is its function in any of the senses mentioned above.

The notion of mechanism, as repeatedly stated in the mechanistic literature, is necessarily connected to the notion of function and often, function just is the phenomenon we are seeking to explain. But are those two things the same? Does saying "mechanism for a function" mean the same as "mechanism of a phenomenon"?

Craver claims that "The world does not come prechunked into mechanisms; it takes considerable effort to carve mechanisms out of the busy and buzzing confusion that constitutes the causal structure of the world" (Craver 2013: 140). The same causal structure can be reconstructed depending on the specific function we are interested in (these claims can be found all over the mechanistic literature, from Kauffman 1971, Bechtel and Richardson 1993, Craver 2001, Craver and Darden 2013 to Craver 2013). For Craver, and for many other mechanists, the decomposition of a mechanism is necessarily tied to a specific function of interest. These functions can reflect the interest of scientists, the scientific community, or a certain research programme but nothing implies that they are there independent of our epistemic interests. Therefore, Craver defines the function of a part of a mechanism and the function of a mechanism itself through its causal role in the overall working of a mechanism or in bringing about a phenomenon rather than using the selected-effects account of functions, fitness account or Boorse's BST account. Since it is a matter of perspective how a system or mechanism is decomposed and what functions are singled out and explained, Craver names this view "perspectivalism".

As mentioned, many mechanists are realists about mechanisms in general, and Craver's perspectivalism (and it seems that Kauffman's view can also be included) seems to be going in a different direction. However, for Craver at least, perspectivalism does not imply agnosticism

or antirealism about real mechanisms.[56] Many mechanisms will share parts. One part can be a part of several different mechanisms depending on the perspective (function) we are considering. For example, cGMP is a component part in the mechanisms of vasodilation in the corpus cavernosum and in light transduction in the retina. Mechanisms can have overlapping parts and activities but that does not mean that they are not real. A clear articulation of this view is perhaps, best given by Glennan: "The fundamental point is that boundary drawing – whether spatial boundaries between parts of mechanisms or between a mechanism and its environment, or temporal boundaries between the start and endpoints of an activity or mechanical process – has an ineliminable perspectival element. But the perspectives from which these boundaries are drawn are not arbitrary or unconstrained. The perspective is given by identifying some phenomenon. This phenomenon is a real and mind-independent feature of the world, and there are real and mind-independent boundaries to be found in the entities and activities that constitute the mechanism responsible for that phenomenon" (Glennan 2017: 44). Craver's perspectivalism is supported by his causal role (CR) account of functions. There, the function of a mechanism or the function of a mechanism's part is relative to the perspective that scientists or the scientific community assumes. The function, then, is analyzed only in terms of the contribution of the component part or mechanism as a whole in the overall "normal" activity of the individualized system: "The sense of 'normal' here is thus not synonymous with 'universal' or 'regular' or 'typical' but instead should be understood as specifying how the [parts] work as they normally do and so on, until the hierarchy ends in some behavior that the scientist is interested, for whatever reason, in explaining" (Craver 2013: 140).

Unlike Craver, Garson (2013, 2018) is explicit in distinguishing functions, phenomena, and their underlying mechanisms. First, he asks several questions that he thinks are from an ontological perspective: "Compare the set of all mechanisms and the set of all mechanisms that serve functions. Are these two sets coextensional? Or is the set of mechanisms that have functions a proper subset of the set of mechanisms that have phenomena? In other words, is it that all mechanisms have phenomena, but for some of those mechanisms, those phenomena happen to be their functions, too?" (Garson 2018: 104, 105). He claims that the function of the

[56] As I noted, even Bechtel, the most ardent defender of the epistemic conception of explanation among the mechanistic philosophers, commits himself to at least some kind of realism about mechanisms in nature in Bechtel and Abrahamsen (2005). Whether this is still his position is uncertain (see footnote 68). Similarly, Ioannidis and Psillos, although taking a certain ontologically agnostic account of mechanisms, do not claim that the causal structures that we try to explain mechanistically do not exist in nature.

heart in the circulatory system is to pump the blood, not to make regular thump-thump noises in a doctor's stethoscope. In his view, then, we can say that the heart is a mechanism *underlying* the thump-thump phenomenon, but not that *this is its function*. Garson uses this to distinguish between minimal and functional mechanisms. Minimal mechanisms are individuated by the phenomenon they produce and underlie, and any functional talk can only be additionally implied or ascribed to them. The heart is a mechanism for both pumping the blood and making thump-thump noises, but the latter is not its function which carries an additional claim of purposiveness or utility. As Glennan says, a volcano is a mechanism for spewing lava, "but there is no hint of design or function in their eruptions" (Glennan 2017: 24). On the other hand, functional mechanisms are necessarily individuated by the function they serve. According to Garson, functional mechanisms are a proper subset of minimal mechanisms. But what kind of functions is Garson talking about?

In his (2018) Garson does not argue for any of the established philosophical positions on function (i.e., the selected effects account – SE, biostatistical account – BST, or causal role account – CR). Rather, his goal is primarily to establish the plurality of talk about mechanisms. Therefore, depending on the definition or theory of functions we prefer, we can talk of SE-functional mechanisms, BST-functional mechanisms, or CR-functional mechanisms. He argues that such functional mechanisms more accurately resemble the way biomedical researchers think and use the notion of a mechanism.

As can be seen from this short reflection, mechanistic philosophers do not quite agree on the notion of function in mechanisms and of mechanisms. Although it has been on the sides of the mechanisms debate, it is an issue that presents a problem when one considers diseases. By examining the consequences of the debate concerning the functions of mechanism Garson, in his (2013) argues that the causal-role account is too permissive to be useful in explaining diseases. He argues that medical and biological sciences take the view that diseases are not due to some disease mechanisms but to the inability of functional mechanisms to perform their functions. There are no disease mechanisms or pathophysiological mechanisms. Diseases, then, are dysfunctional or broken physiological mechanisms. For now, I will only raise this worry and will return to it in the final two sections of this chapter where I will consider mechanisms of diseases in particular.

What about phenomena? How should we characterize phenomena? Craver and Darden write: "Phenomena are typically the kinds of things that potentially can be detected,

manipulated, or produced in many ways across different experimental arrangements or observed with a wide variety of observational methods" (Craver and Darden 2013: 55). Therefore, the characterization of a phenomenon is inevitably shaped by the "accepted or available experimental protocols for producing, manipulating, and detecting it" (Craver and Darden 2013: 55). But a phenomenon is not a collection of data. Data is what we gather through experiments and observations. Data only imply the existence of a phenomenon; they do not constitute it (see Bogen and Woodward (1988) for more on this view). Weber in his (2009) argues for a similar view and asserts that an explanandum is constructed out of data, not out of a phenomenon. If a doctor gathers information about your symptoms and signs and later runs certain diagnostic tests, the information gathered is not the phenomenon and does not represent the phenomenon. It is only indicative that there is a mechanism, or in this case, that there is a certain mechanism of disease present in the body. Mechanists usually discuss repeatable phenomena, such as protein synthesis, long-term potentiation, or light transduction in the eye. But a phenomenon, at least as it is discussed in the mechanistic literature, need not be a repeatable event.

The importance of a particular definition of the phenomenon one wants to explain cannot be overestimated, whether we are discussing ontological, epistemological, or methodological mechanisms. Remember from the introductory section of this chapter that mechanisms are always defined as mechanisms for or of some phenomenon. Epistemologically and methodologically speaking, the characterization of a phenomenon presents a first step in developing a model of mechanism. It has continuing consequences on the whole project of developing a model. But for mechanists, constructing a mechanistic explanation also influences how, in the end, the phenomenon itself is characterized. Illari and Williamson's quote reflects the same ideas found in Glennan's, Bechtel's, Craver and Darden's work: "Mechanisms are individuated by their phenomena, and phenomena are also individuated by their mechanisms. This is not circular, because it happens iteratively over time. At the beginning, a mechanism is not needed to individuate a phenomenon, but the characterization of the phenomenon may be further refined when a mechanism or mechanisms are discovered" (Illari and Williamson 2012: 124).

The way that scientists characterize a phenomenon (or a function) inevitably influences, navigates, and steers the investigation process. How? It provides, as Darden succinctly puts it, "guidance and constraints" (Darden 2013: 20). Recall that many mechanists take that the

characterization of a phenomenon influences the particular decomposition into parts and activities needed for a phenomenon to occur. This characterization then defines the space of possible mechanisms able to produce the phenomenon so characterized. Finally, in addition to constraining the space of possible mechanisms responsible for it, the characterization of a phenomenon influences where and how we search for the mechanism and its boundaries. Subsequent investigation usually follows two general steps. First, as stated previously, the characterization of a phenomenon limits the space of possible mechanisms. As Craver and Darden claim, rarely do explanations of biological mechanisms start from scratch, or from complete ignorance. Scientists observe similarities between certain phenomena which indicate that previously discovered mechanisms, entities, or activities might be used to explain them. For example, knowing the mechanisms of action of cAMP proved to be very useful in illuminating the intricacies of the NO-cGMP causal pathway. Here, Craver and Darden's considerations are in line with Ioannidis and Psillos's ideas about the language used in the process of discovery and explanation construction. They write: "To describe a phenomenon is to characterize it in the language of a given field and to implicitly call up the host of explanatory concepts, the store of entities, activities, and organizational structures known to a field at a time, that might be used to construct a schema of the mechanism" (Craver and Darden 2013: 52). Second, mechanistic explanation has to consist of three features: parts, activities, and their overall organization. In that regard, Bechtel and Richardson's book *Discovering Complexity* is perhaps the first and most comprehensive account of this aspect of the mechanistic methodology in the literature. There, they present two strategies they think are the most useful and widespread heuristics of the mechanistic methodology: decomposition and localization. Both decomposition and localization follow from and are directly influenced in their implementation by the particular characterization of a phenomenon.

I mentioned in section 2.2. that decomposition refers to the strategy of breaking down a system into its constitutive parts. Decomposition, as Bechtel and Richardson claim in (1993), can be achieved in two ways – structurally or functionally. Functional decomposition, as its name implies, is similar to the functional analysis of a system's behavior, for example in Cummins (1975) and (2000). Cummins's functional analysis is intended to explain a behavior or capacity of system by decomposing this capacity into subcapacities and showing how these subcapacities work together so that they produce or underlie the overall capacity of a system. Similarly, functional decomposition refers to identifying and locating lower-level operations which contribute to the overall functioning of a mechanism. Structural decomposition refers to

the decomposition of a system into its active parts or component parts. Identification of working parts of a mechanism can be a starting point, as Bechtel and Richardson and Bechtel and Abrahamsen claim, but it is not necessary that we immediately know which ones are indeed the working parts of a mechanism and what their causal roles are in detail. A proposition about the parts that might compose a mechanism comes from the characterization of a phenomenon and knowledge about properties of various entities. This, then, offers some grounds for hypothesizing that these entities are present in the mechanism of our interest. The next step is localization – the assignment of causal roles to these already identified parts. This is the process that connects the proposed lower-level activities and operations with already identified component parts. An inability to link an operation to a part leads to doubts about whether one of them (or both) is really present in a mechanism.

All three theses of mechanistic philosophy consider mechanisms as having layered nature. Mechanisms have parts which are also sometimes mechanisms. That is, parts can always be decomposed into parts and so on (until we reach the level considered to be fundamental for a given group of scientists, as MDC stated in their 2000). How do we know which entities are constitutive parts of mechanisms? As I already noted, the rationale in discovering the organizational and structural decomposition of etiological or horizontal mechanisms is best described by Woodward's interventionist account of causation. Many mechanists adhere to this and there is not much controversy concerning it (if at all). After all, Woodward himself acknowledges to construct his account by looking closely at scientific practice. But how scientists discover the constitutive relations between parts has been a subject of much controversy. Craver first proposed his account of "mutual manipulability" which was also based on the interventionist account (for example, in his 2007a). Since its inception the mutual manipulability account has sparked a lot of criticism, mainly focusing on the claim that it confuses causal and constitutive (part-whole) relations (e.g., Leuridan 2012, Baumgartner and Gebharter 2016, Kästner and Andersen 2018). There have been several attempts to resolve the issue, coming from both a metaphysical and epistemological point of view (e.g., Craver and Bechtel 2007, Gebharter 2017, Baumgartner and Casini 2017). Interestingly, not all of the critics of the mutual manipulability account argue against constitutive mechanisms and constitutive relevance. Rather, the point of contention is what is the correct metaphysical characterization of it and what kind of strategy should we use for constitutive inferences (for the latest proposed approach, consider the "matched interlevel experiments" account developed in Craver, Glennan and Povich 2021).

The previously mentioned four stages of mechanistic investigation are constantly revised by scientists as their research progresses. The characterization of a phenomenon, although being the first step, is constantly shaped by new discoveries about the mechanism responsible for it – that is, by the construction and evaluation of the mechanism schema. The characterization of a phenomenon influences how the mechanistic investigation and the construction of an explanation proceeds, but as the investigation reveals new features, the mechanism schema makes us revise the characterization of the phenomenon. The four stages of mechanistic inquiry are always in a certain interplay with each other. This characteristic of mechanistic methodology is best described as an iterative process where the mechanistic model and the predictions based on it are constantly being revised (Darden 2006, Craver and Darden 2013, Becthel, Abrahamsen and Sheredos 2018). I will discuss this process in a more detail in the next chapter, where I will be specifically concerned with mechanistic predictions and how they figure in mechanistic investigative strategies.

2.6. The relation between ontological mechanisms and their models

In the previous three sections, I presented my disambiguation of mechanistic philosophy into three connected yet clearly separated theses and touched upon numerous issues surrounding mechanisms (and related notions, such as "mechanistic" or "mechanical"). Nonetheless, I only brought these issues to attention and mainly refrained going into deeper discussion. In the following sections, I provide resolutions to some of these issues and establish my position on the limits of the OM thesis about mechanisms and discuss criteria of the explanatory power of models of mechanisms. The conclusions will be important for the discussions of the next chapter, where we undertake the analysis of the recurrent failure of mechanistic reasoning to provide true predictions in medicine and discuss a set of criteria for good mechanistic reasoning. Of particular interest here is the question of the relation between models of mechanisms and the causal structures in the world or portions of the world that they are supposed to represent. Therefore, I identify two clusters of questions, mainly corresponding to the ontology and epistemology of mechanistic philosophy. First, I assess ontic mechanisms: what are ontological mechanisms supposed to be and how plausible (that is, metaphysically serious) are the accounts provided in the literature? Second, how are epistemic mechanisms connected to ontological mechanisms and what can be said about the ontic and epistemic conceptions of mechanistic explanation in regard to the first question?

2.6.1. Ontological mechanisms and the ontic versus epistemic conceptions of explanation

Recall how Salmon argued that in the CL account (specifically, the deductive-nomological form of scientific explanation), an explanation is informative (and, in the end, true) because of its logical structure – states of affairs, facts of the matter or initial conditions are subsumed under a law of nature of general scope. Hence, Salmon defines it as an epistemic kind of explanation – the explanation does include a metaphysical or ontological thesis, namely laws of nature, but it is, after all, a matter of whether the conclusion follows from the premises. On the other hand, Salmon argues that scientific explanations do not work like that. Salmon argues that scientific explanations are usually concerned with revealing or bringing to light the underlying causes which bring about phenomena. Therefore, the principal aim of scientific inquiry towards constructing an explanation is finding out the causal structure of the world. In that regard, Salmon writes: "The relationships that exist in the world and provide the basis for scientific explanations are causal relations" (1984: 121). Scientific explanation, then, is not explanatory because it is a well-formed deductive argument, but rather it is informative and explanatory because it cites causal processes in the world found by our best empirical sciences. As he has noted, his account "is an attempt to put 'cause' back in the 'because'" (Salmon 1977:215).

Notice, then, how in Salmon's conception of scientific explanation, causal relations in the world, the world's causal structure or architecture, just is the explanans, while phenomena or the effects of causal relations are the explanandum. In a nutshell, a cause explains its effect by causing it. Since scientific explanations are explanatory because they cite ontological constituents, he takes this conception of explanation as an ontic conception of explanation.

Recall Hume's classic example of colliding billiard balls. In his early definition of causation, Salmon defined an object as a causal process if it is capable of transferring a mark and carrying it on after the transfer. What does this mean? Salmon claimed that one billiard ball starts to move upon collision with another billiard ball, not because it is a regularity observed in numerous instances in the past and in other locations, but rather because of the transfer of the mark – in this case, momentum – between the balls. The ball which acquired the mark is capable of moving because it is capable of retaining that mark along its world line. Salmon uses this theory of causation for his ontic conception of scientific explanation. The

explanation reveals the connection between the purported cause and its effect – transfer of momentum.[57]

Later mechanistic philosophers have taken the ontic/epistemic terminology and applied it to a distinction pertaining to explanatory force in the mechanistic account of explanation. On the one side, proponents of the ontic conception claim that the mechanism responsible for the phenomenon explains the phenomenon. Recall how Craver distinguishes between objective explanations and explanatory texts. In this case, an objective explanation just is the mechanism itself. Similar to Salmon's claims, by being constitutive of a phenomenon or causally responsible for it, the mechanism underlying a phenomenon explains that phenomenon. These claims are repeated in similar fashion in, for example, Illari and Williamson (2011) although they call them physical explanations. As they claim, "in the epistemic sense of explanation it is the *description* of the mechanism that explains, while in the physical sense, the *mechanism itself* does the explaining" (Illari and Williamson 2011: 822). Such an idea, however, comes directly from Salmon's conception of scientific explanation. By locating the occurrence of a phenomenon within the causal structure of the world, we explain the phenomenon. In the mechanistic take on the ontic conception of explanation, mechanisms explain the phenomenon and it is the business of the sciences to find out those mechanisms and their features and characteristics. On the other side, proponents of the epistemic conception claim that scientific explanations (that is, models of mechanisms or mechanistic explanations) are representations of these physical entities (whatever they are). Models have different non-ontic, epistemic features which help us to "epistemically grasp" the phenomenon. Mechanism by itself does not explain anything. How could an inactive mechanism explain the phenomenon which it is supposed to be responsible for?[58] Therefore, scientific explanation is an epistemic endeavor. It

[57] At that time Salmon took causal relations as being mark transferals. What explains the movement of the billiard ball then, is the transfer of momentum which itself is the connection between two causal processes. We have explained the movement of a ball by citing the cause(s) of the movement – transfer of momentum from one ball to another upon collision. Under the influence of Phil Dowe's effective criticism (e.g., in Dowe 1992)), Salmon later abandoned the idea of mark transfer and accepted Dowe's idea of causation as the exchange of conserved quantities (e.g., momentum or charge) but the foundations of his account of explanation remained the same. Such a theory of causation later became known as the Salmon-Dowe theory.

[58] As a connecting issue, Craver (a proponent of the ontic conception) argues that there could be a mechanism that we will never find out about or a mechanism that we will never be able to understand but, as he claims, it could still, as a matter of being causally or constitutively responsible for phenomena, explain those phenomena.

is a matter of making things in the world *understandable to us.* As Wright and Bechtel claim: "After all, explaining refers to a ratiocinative practice governed by certain norms that cognizers engage in to make the world more intelligible; the non-cognizant world does not itself so engage" (Wright and Bechtel 2007: 51). Abstraction, idealization and, equally important, generalization, are epistemic characteristics of scientific explanations, and these cannot be found in ontological mechanisms even though explanations of mechanisms could, in fact, be true of their target systems.[59]

As I have repeatedly noted, Salmon's ontic conception of explanation has been very influential for the proponents of the ontic conception of mechanistic explanation (in Craver 2007a, for example). Critics of the ontic conception have identified several problematic consequences of such a conception of scientific explanation. Some of them have been addressed, while others have been ignored. For example, it has been argued that the ontic conception of explanation implies (or maybe better, necessitates) that scientific explanations are token event explanations. But as Wright and Van Eck argue quite convincingly "scientific explanations of phenomena are not case studies in tokening" (Wright and van Eck 2018: 1017). Although not completely true (for example, explanations in the historical sciences seem to be largely of token events rather than types of events), it is certainly correct that most scientific explanations are concerned with types of phenomena rather than particular events. Biological sciences predominantly look for explanations of types of events. For example, mechanistic explanations of protein synthesis or the relaxation of smooth muscle cells do not address the synthesis of some specific protein molecule at a specific place and time. Explanations of medical phenomena are not different in that respect. All representations of the heart in medical textbooks do not represent a specific heart at a specific time but rather a type of a mechanism in a very abstracted and idealized way. The textbook model of the heart, quite possibly, does not exactly correspond to any actual organ. Even when there is a photograph of a specific heart in a medical textbook, the purpose of that representation is not to give a token explanation of *that* heart. A medical explanation of cardiac arrest does not refer to *the* cardiac arrest of patient *X* from a country *Y* at a time *T*. Wright and Eck express this view more vividly: "while excavating token cadavers has played an important pedagogical role in the education of legions of medical students, the scientific explanation of heart disease is not an ontic explanation

[59] For more on the discussion on ontic/epistemic distinction in mechanistic literature, consider Craver (2007a), (2014), Wright (2012), (2015), Illari (2013), van Eck (2015), Wright and van Eck (2018).

residing in a single token chest cavity" (Wright and Van Eck 2018: 1017). Although a nuisance for the account, defenders of the ontic conception can try to accommodate this problem by generalizing over single causal relations. A textbook model of the heart is an explanatory text rather than an objective explanation, but it is an idealization and abstraction of numerous objective explanations. Although this is an issue for proponents of the ontic conception of mechanistic explanation, I will not elaborate further on it. I wish to discuss a rather different problem for the ontic conception of mechanistic explanation. I claim that models of mechanisms sometimes require features which are problematic to understand ontologically but, nevertheless, seem necessary for understanding and explaining mechanisms.

The ontic conception of explanation, as mentioned above, requires that each objective explanation (in Craver's terminology) is a token-event explanation. Furthermore, it requires that mechanisms are particulars, where the causal relations between their parts are singular causal relations, preferably, or even necessarily, involving active rather than passive metaphysics, and actual rather than counterfactual notions.[60] But as I argue, actual and active causation (whatever the theory of causation along those lines we consider) is not (entirely) compatible with the claims of the OM thesis and the ontic conception of mechanistic explanation. Why?

First, notice that nothing said so far makes mechanisms distinct causal structures from, for example, Salmon's causal network. Indeed, Ioannidis and Psillos have argued that generative mechanisms, as they call them, do not presuppose something distinctively different other than "any relatively stable arrangement of entities such that, by engaging in certain interactions, a function is performed, or an effect is brought about" (Ioannidis and Psillos 2018: 147). Further ontological claims have to be brought in in order to make mechanisms those things that mechanistic philosophers have implicitly been assuming. It does not matter whether we take causation as an unanalyzable primitive relation, manifestations of causal powers, or an exchange of fundamental properties, entities engaging in activities or interacting with one another are constitutive of different general metaphysical theories of causation irrespective of the reality of mechanisms (e.g., Salmon 1998, Cartwright 2007, Mumford and Anjum 2011). On the other hand, probabilistic, interventionist, and other counterfactual accounts of causation

[60] Consider Glennan (2011) for the argument that mechanisms are particulars, and Illari and Williamson (2011) for the argument that mechanisms require active metaphysics.

included in, for example, models of causal Bayesian nets, structural equations, and other structures representing networks of (causally) interrelated variables do not necessarily come equipped with something related to the mechanistic ontology discussed in the OM thesis. However, and in spite of all that has been said, mechanisms do possess something that is, according to mechanistic philosophers, able to distinguish them from merely being a relatively stable arrangement of parts or Salmon's world-encompassing causal network. As I presented at the beginning of the chapter, mechanisms are not aggregative structures. They possess numerous relations of different kinds which makes them something other than simple aggregative structures. Therefore, what makes something a mechanism is that the activities and interactions of the entities that constitute them are organized in some specific way so that a phenomenon is produced *because* of that organization. Furthermore, in many types of mechanisms, it is just that specific organization that makes a specific mechanism just that kind of a mechanism.

Mechanism might be internally organized in different ways. Sometimes their organization can consist in simple spatial and temporal relations between their parts and sometimes there might be more complex causal relations between them, such as feedforward and feedback signals. Models of mechanisms can represent them quite efficiently. Diagrams and visual models, such as the ones in Figure 7 and Figure 8, can represent spatial relations easily. Using arrows and other features, these models can represent temporal relations too. Different models can represent the causal organization within a mechanism (for example, in the form of sets of equations). But some organizational features cannot be represented as properties or relational properties between parts but rather as abstract features or mathematical relations between parts and activities (and interactions). If this is true for some mechanisms, then the ontic conception of explanation raises then the question of the ontological status of such organizational features: how are we supposed to relate such features of mechanistic models to ontological mechanisms? I argue that to account for the regularity and maintenance of a phenomenon which a mechanism is responsible for, organization sometimes plays a specific causal role which can only be accounted for by some difference-making causal relation. I present an argument from my (2021) where I discuss how this causal role of organization in mechanisms can be accounted for. There, I argue that both kinds of concepts of causation (production/mechanistic and difference-making) are needed in models of mechanisms as well as in ontological mechanisms. But then this, then, I claim, has consequences for anyone wishing to uphold claims that ontological mechanisms are real and

local, as Illari and Williamson argue (i.e., mechanisms require only active metaphysics which precludes any non-local and counterfactual features), and that mechanistic explanation should be understood along the lines of the ontic conception of explanation (where there is no place for abstracta, genera and idealizations in ontological mechanisms).

2.6.2 A problem for ontological mechanisms and the ontic conception of explanation

To help clarify my argument, I use a well-known phenomenon from molecular biology, already discussed in my (2021) – the operation and maintenance of the genetic switch in bacteriophage lambda. The case study at hand has become a paradigmatic case for studying developmental genetic networks (Oppenheim et al. 2005), such as, predictive computational modeling of changes in the expression of genes (Vohradsky 2017). Because of its rather simple functioning, the genetic switch in bacteriophage lambda provides an excellent introduction and preview of the complexities of genetic mechanisms and an example of how organizational features causally matter in the maintenance and production of a phenomenon.

Let me start with a brief presentation of the phenomenon and its underlying mechanism before discussing what exactly I found to be the problem for the ontic conception of mechanistic explanation and Illari and Williamson's argument that mechanisms ought to be real and local.

Bacteriophage lambda is a virus that infects *E. coli* bacteria. It consists of a single DNA molecule that is surrounded by a protein coat. After attaching onto the surface of an *E. coli* cell, the phage particle drills a hole in the surface and inserts its chromosome into the cell. When the viral chromosome has been inserted, phage lambda can enter into two different modes of replication. These two modes are dependent on a number of conditions. On the one hand, if the environment of the *E. coli* cell is rich in nutrients, the phage enters the lytic cycle or replication. This is a mode of extensive replication of the phage chromosome. After a period of approximately 45 minutes, the bacterium infected by the virus "lyses" and about 100 new progeny phages kill the bacterium by bursting into the environment. In a different mode of replication, which happens in most cases of infection, the phage enters a lysogenic mode. In this mode, the viral chromosome (now called prophage) integrates itself into the host chromosome. Here, however, the prophage is passively replicated along with the bacterium's

own replication. This mode of replication can now be prolonged indefinitely. But, conditional on a stimulus from the environment (for example, the irradiation of the bacterium by ultraviolet light), the phage will undergo a process called lysogenic induction. The mode of replication then switches from lysogeny to the lytic replication cycle.[61]

The mechanism that changes and maintains the mode of replication consists of the following entities. Two genes determine the passive (lysogeny) and active (lytic) modes of replication. Each of the two genes has its own promoter, that is, a site which points RNA polymerase in different directions. The P_{RM} promoter points polymerase leftward while the P_R promoter points polymerase rightward so that it starts to transcribe one or the other gene. The activities of these promoters are regulated by the binding of two different kinds of proteins onto three operator sites - O_{R1}, O_{R2}, and O_{R3}, which then determine the mode of replication. In the lysogenic mode, at least two of the repressor proteins called "cI" have to bind simultaneously onto the operators O_{R1} and O_{R2}. This happens often since they usually bind cooperatively. The binding of one repressor protein onto the operator site O_{R1} prevents the binding of RNA polymerase to the right promoter, and thereby disables the expression of the gene responsible for the lytic growth mode (C*ro*). The binding of the other repressor cI onto the O_{R2} or O_{R3} sites activates the expression of the gene *cI*, responsible for the lysogenic growth.[62] On the other hand, only one repressor protein is enough for the lytic replication mode. That is, the lytic replication mode requires a negative regulation. The O_{R3} operator has a low affinity for the cI proteins. Here, the expression of gene *Cro* is activated when the repressor protein "Cro" binds to the operator O_{R3} and by that silences gene *cI*.

This mechanism has additional important characteristics in order to function as it does. Both kinds of repressor proteins regularly bind and fall off, irrespective of the particular mode of replication. That is, a single protein does not stay attached to the operator site throughout the specific mode of replication. A different kind of protein might come to be attached at some time yet the same mode of replication continues. Why does this happen? The mechanism of the genetic switch of phage lambda is sensitive to a very precise ratio of concentrations of both kinds of repressor proteins. If the concentration ratio of cI to Cro repressor proteins satisfies a certain threshold point (depending which mode of replication is activated), either *cI* or *Cro*

[61] For a detailed description see Ptashne (2004).
[62] The gene names are italicized while their corresponding protein names are not.

gene continues to be expressed, regardless of the situation where the "wrong" protein is at the moment bound onto the operator(s).

To understand the sensitivity of the mechanism to the concentration ratio of repressor proteins, consider this passage from Ptashne (2004):

> (…) in a lysogen enough repressor is synthesized to repress P_R about 1000-fold. Over the first twofold or threefold drop in repressor concentration from this high level, the activity of P_R remains unchanged. In effect, repression is buffered against ordinary fluctuations in repressor concentration, so that lysogens are rarely "accidentally" induced. But when the repressor concentration has dropped about fivefold, P_R responds dramatically, functioning at about 50% of its fully unrepressed level. This allows synthesis of Cro and of other lytic gene products, thereby flipping the switch.
>
> Ptashne 2004: 26

The regulation of the mode of replication seems to be directly dependent on the concentration ratio between populations of different proteins. Therefore, the following problem emerges: considering the mechanistic explanation of this phenomenon, how should we understand this relation of dependence on the concentration ratio itself?

My argument consists of two parts. First, I show that active causation cannot account for the stability of mode of replication in cases where the "wrong" protein is bound to the operator sites. Second, to grasp the phenomenon and clearly distinguish causal roles in the mechanism, I argue, both mechanistic and difference-making approaches to causation are needed.

In (2021), I have argued that theories of causation that consider causation as a change in the properties of objects (or entities) cannot explain this stability of the mechanism of gene expression. Such accounts can only provide "a snapshot": "a description of all the physical interactions between molecules relevant to a given explanation, occurring at time *t*" (Nathan 2014: 200). But the dependence relation between the stability of the switch and the concentration ratio between the different repressor proteins do not involve an activity or a change of a property of any working part of the mechanism (operators, proteins, RNA etc).[63]

[63] See Nathan (2014) pp. 199-200, for the complete argument against the availability of the Salmon/Dowe process theory to account for the stability of the genetic switch. The "snapshot" argument applies equally to the MDC approach.

On the other hand, Woodward's interventionism can explain the stability and change of the modes of replication, but it has other limitations. In principle, by satisfying all of Woodward's criteria, interventions that set the concentration ratio cI:Cro to different values would change the functioning of the genetic switch. Nevertheless, the interventionist account falls short of explaining the difference in causal roles of the concentration ratio of proteins and of protein-operator interactions. That is, the problem is that Woodward's interventionist account cannot acknowledge or distinguish the causal roles of actual causes – the proteins bound to the operator sites – and the back-up causes which constitute the concentrations of proteins and hence their concentration ratio.

Considering the OM thesis of mechanistic philosophy and the ontic conception of explanation a familiar problem reappears. How should we understand these back-up causes given these theses? Furthermore, an even more pressing matter is the metaphysical poverty of Woodward's interventionist account of causation. The interventionist account lacks an accompanying metaphysics to tell us what the variables we choose to intervene upon and measure the values of stand for. What are the limitations on choosing the variables we want to measure, concerning the ontological category they ought to represent? How should we interpret these variables from the perspective of the ontic conception of explanation? Woodward's interventionism can recognize and confirm the causal relevance of the concentration ratio, but it remains silent on how and why we think it differs from the causal role played by the activities and interaction of particular proteins.

Now, why do I consider that to explain this mechanism, we have to acknowledge the irreducible causal role of the concentration ratio, different from actual/physical/mechanical causal relations, expressed in for example, the protein-operator interactions (no matter how complex and numerous the set of those interactions may be)? Recall the discussion on what kind of things a mechanism's component parts are supposed to be. Component parts are entities or things which engage in activities and interactions by being the bearers of causal powers, they are spatiotemporally located, structured, and oriented, and usually they are also decomposable into parts. However, mechanistic philosophers have claimed that this does not imply that entities cannot be dispersed across some regions of space, nor that a population of entities cannot be considered as a working part of a mechanism in a mechanistic explanation. Indeed, a lot of mechanistic explanations across the sciences look like this. Often, parts in models of mechanisms are just given as a type of an entity, without literally specifying every

single particular of that type. For example, in the mechanism of smooth muscle cell relaxation, the population of cGMP molecules is considered as a working part of that mechanism, and not individual cGMP molecules themselves. If we apply this to the case at hand, we can give a mechanistic explanation of the genetic switch phenomenon where the populations of proteins cI and Cro, rather than individual proteins composing these populations, are working parts. That is, in this case, the concentrations of proteins are properties of working parts of a mechanism. By being the bearers of such properties, they have causal powers needed for the overall working of the mechanism. Could the argument be stretched even further? Could we claim that even the whole populations of both cI and Cro proteins combined constitute a single working part of the mechanism, thereby making the concentration ratio itself a property or a feature of that working part? That is, could we try to construct a mechanistic explanation of this phenomenon where the overall population of both proteins cI and Cro is a working or a component part of the mechanism of the genetic switch of phage lambda? The concentration ratio, then, would be a property of one of the component parts of the mechanism.

However, this cannot be achieved. Why? First, they have different structures. The cI protein is a dimer: it is composed of 236 amino acids folded into two domains connected by a string of 40 amino acids. The Cro protein, on the other hand, is a monomer. It is composed of 66 amino acids.[64] Second, these proteins possess different causal roles within the mechanism as is nicely recognized in the relevant literature on the phenomenon: "These two proteins – repressor and Cro – bind to the same three operator sites but play opposing roles in the switch mechanism" (Ptashne 2004: 16). Since these populations of proteins have different structures and occupy different causal roles, they are different component parts in the explanation of the mechanism of the genetic switch. Therefore, if these populations are different component parts, then the concentration ratio is not a property of a single component part, nor it is a property of either one of the populations of cI and Cro proteins. Finally, since it is not a working part, and it is not a property of some part of the mechanism, what is the concentration ratio? The concentration ratio is an abstract mathematical relation between parts of the mechanism of the genetic switch. The only way to understand this within the context of mechanistic explanation, as I claimed in (2021), is it to define the concentration ratio as a kind of organizational feature, in the same sense as the temporal or spatial relations of a mechanism's parts.

[64] Although, to be more accurate, they almost exclusively form dimers since the affinity of Cro monomers for each other is high.

How do organizational features matter causally? Consider an argument given by Levy and Bechtel in their (2013). They argue that the specification of details concerning parts, their relations and in virtue of what properties they fulfil their causal roles does not always a provide an optimal explanation of a mechanism. They write: "It is always possible and, we argue, often desirable to overlook the more concrete aspects of a system and represent its organization abstractly as a set of interconnections among its elements. Oftentimes such a detail-poor representation will be well suited for the explanatory purposes at hand" (2013: 255). To explain most if not all complex mechanisms, these abstract organizational features are necessary. Levy and Bechtel's claim that ″altering the details of the components (as long as they meet the minimum conditions for fulfilling the role in the organizational schema) does not change the behavior, whereas altering the organization (changing what is connected to what) does″ does not just amount to a strategic principle but rather the necessary characteristic of most or maybe all models of complex mechanisms (2013: 253). Organizational features (e.g., feedback and feedforward loops, frequencies of a signal, the sensitivity of a genetic switch to the concentration ratio of different proteins) represent boundaries within which mechanisms works properly. That is, they are difference-makers for bringing about phenomena. For example, the mechanism of the genetic switch in phage lambda could still function properly if the properties of proteins were somehow slightly changed (e.g., their binding affinities being slightly lower or higher than the actual case) but changing the organizational feature of sensitivity of the genetic switch to precise levels of the concentration ratio would significantly alter its behavior.

In this case, the individual entities or component parts of the genetic switch mechanism are proteins, operator sites, promoters, and RNA polymerase. They are the real, robust, concrete, and material objects or continuants that engage in causally productive relations. The concentration ratio and the distribution of proteins are features of the organization of this particular mechanism. Neither of the two causal relations is sufficient, but both are necessary to explain how the replication cycles of the virus function. Therefore, the causal role of proteins and operators and the causal role of the concentration ratio between proteins populations are both needed for the phenomenon to occur, but they are responsible for two completely different aspects of the phenomenon. Proteins produce the flipping of the switch and gene transcription (an actual, physical, or mechanical causation), but the stability, regulation, or maintenance of the gene transcription once the switch has been activated counterfactually depends on the concentration ratio of proteins. In other words, it is an organizational feature of the mechanism that determines the relevant counterfactuals regarding the output of the mechanism of gene

expression. In his (2014), Nathan goes even further. He argues that this example shows that there is a specific kind of counterfactual causal relation which he calls *causation by concentration*. Whether we need to introduce a new kind of counterfactual relation to explain this or not, it does not change the fact that the relevant threshold of the concentration ratio in this case is indeed a difference-maker.

The problem for the ontic conception of mechanistic explanation and ontological mechanisms should be clear by now. Complex mechanisms have complex organization. Mechanistic explanations of many of these complex mechanisms will have to account for the difference-making causal roles of the organizational features in the production of the phenomenon which (as we have seen in the case of phage lambda) cannot be explained by actual, physical causation. But counterfactual, difference-making relations cannot be constitutive of ontological mechanisms, or so I claim. The problem remains even if counterfactuals are understood along the lines of Woodward's interventionist account, rather than Lewis's counterfactuals. Illari and Williamson recognize this themselves: "Nevertheless, if you did take Woodward's position to be a metaphysical one, his invariance relations would be non-local, albeit more local than either a modal realist or best-system laws view. Invariance relations would still depend on what happened elsewhere and at other times in this world" (Illari and Williamson 2011: 835, 836). Indeed, Woodward's theory is not always taken to have a metaphysical background (let alone one as strong as Lewis's theory). It is often viewed as an epistemological and methodological account of causation. But it is, nonetheless, a counterfactual theory of causation. In the end, Woodward himself admits this: "[Difference-making] causal claims involve a comparison of some kind between what happens in a situation in which a cause is present and alternative situations (which may be actual or merely possible) in which the cause is absent or different" (Woodward 2011, 411). But how can this work for real particulars instantiated at some space and time?

Most mechanistic philosophers have claimed that "The New Mechanistic Philosophy" is committed to scientific realism. In addition to the methodological turn and the recognition of the centrality of the notion of mechanism to the practice and explanations of the life sciences, Wimsatt (2017) sees the commitment to scientific realism as another feature of the mechanistic philosophy. Indeed, as I claimed at the beginning of this chapter, most of mechanistic philosophers commit themselves to realism about mechanisms and would uphold the OM thesis. So, what does this realism amount to? In this section, I have presented a problem for

mechanistic philosophers whose realism is defined by the OM thesis and the ontic conception of explanation. The EM thesis, as I show in the following section, does not necessarily force one to accept strong scientific realism imposed by the OM thesis. It is enough to uphold what Glennan calls "a minimal form of scientific realism". That is, it is enough to have a stance which "supposes that mechanisms and their constituents are things in the world that exist independently of the models we make of them" (Glennan 2017: 10). This is the stance that seems most fitting for at least two reasons. First, the ontic conception of explanation, as I show, has problems incorporating different epistemic features of models, such as abstract relations, difference-making relations between these features and the phenomenon, and the type-token relations between real mechanisms and models of mechanisms. Mechanistic explanations describe features and portions of the world, but we should not forget that, in the end, these are just models, full of epistemic features, properties and heuristics. Second, the methodological turn in philosophy of science brought about by mechanistic philosophy took scientific practice as a principal guide. But I have also shown that the OM thesis does not come from scientific practice. What will change in scientific practice if we disregard ontological mechanisms and the ontic conception of mechanistic explanation as a philosophically not well supported theses? As Ioannidis and Psillos observe: "What is added to scientific practice by insisting that a description of a mechanism has to be couched in some preferred philosophical categories, e.g., entities and activities, powers, or what not?" (Ioannidis and Psillos 2018: 5). In a nutshell, my claim is that the EM thesis does not require realism about mechanisms in nature if these are understood along the lines of the OM thesis.

2.6.3. What is a good model of mechanism?

If the conclusion from the previous section holds, then, how do we know when a model of a mechanism is a good, complete, or true model of whatever it describes?

Craver and Kaplan are usually singled out as the most vocal proponents of the view that details presumably make an explanation of a mechanism better or more complete. Let me use their abbreviation MDB for the "more details are better" view. The MDB view supposedly goes contrary to much of the practice of modeling mechanisms. Abstractions and idealizations are features of every model. Furthermore, abstraction and idealization are not only present in the practice of modeling, but they even seem necessary for understanding how a mechanism

works (e.g., in Levy and Bechtel 2013 or Love and Nathan 2015). Craver and Kaplan, of course, recognize that modeling serves the purpose of understanding (some aspect of a phenomenon) and that the need for more details does not always (and perhaps in most cases does not) satisfy that purpose. The complete description of a mechanism, involving all the details of its parts, activities, interactions, and organization is too strict a criterion. Craver in his (2006) is quite explicit about this. He writes: "Few if any mechanistic models provide ideally complete descriptions of a mechanism. In fact, such descriptions would include so many potential factors that they would be unwieldy for the purposes of prediction and control and utterly unilluminating to human beings" (Craver 2006: 360). Imagine just how much detail a description of a protein synthesis mechanism would have to be included in order to be a complete description (in the literal sense of the term). These claims are repeated in Kaplan and Craver (2011): "the idea of an ideally complete how-actually model, one that includes all of the relevant causes and components in a given mechanism, no matter how remote, negligible, or tiny, without abstraction or idealization, is a philosopher's fiction. Science would be strikingly inefficient and useless both for human understanding and for practical application if it dealt in such painstaking minutiae" (Kaplan and Craver 2011: 609, 610). Similarly, Craver in his other paper writes: "The ontic explanatory structures are in many cases too complex, reticulate, and laden with obfuscating detail to be communicated directly" (Craver 2014: 28). The search for completeness rarely describes the process of modeling. In that regard, Kaplan and Craver continue, "Yet these commonplace facts about the structure of science should not lead one to dispense with the idea that models can more or less accurately represent features of the mechanism […] and that models that describe more of the *relevant features* of the mechanism are more complete than those that omit them" (Kaplan and Craver 2011: 609–10, emphasis added). But then, what should be the criterion or the criteria for grading the "goodness" or completeness of a mechanistic explanation? That is, what makes some features of a mechanism more relevant for explanation?

In their (2020), Craver and Kaplan propose an answer. First, they provide the background for assessing completeness which comes directly from their acceptance of the ontic conception of explanation. Causal relations in the world are explanatory, and it is the business of the sciences to reveal their existence. In that sense, "the causal structure of the world defines the limit of completeness in one's explanatory knowledge" (Craver and Kaplan 2020: 300). They label this notion of completeness as "Salmon-completeness" (SC) and they define it as following: "The Salmon-complete constitutive mechanism for [the phenomenon] P versus P'

is the set of all and only the factors constitutively relevant for P versus P'" (Craver and Kaplan 2020: 300). Here, they use the contrastive aspect of explanation. A model or an explanation is supposed to explain why or how P happened, rather than P'. The details that are relevant and which a good model of mechanism must include are the details that make a difference to whether P occurred rather than P'. Different explanandum contrasts, they say, are defined or described by different "switch-points": the contrast between being at 0° C rather than above 0° C explains why water is frozen while that between -5° C rather than -10° C does not. It is a difference that makes an explanatory difference, so to speak. Next, they introduce the "More Relevant Details The Better" criterion: "If model M contains more explanatory relevant details than M* about the SC mechanism for P versus P´, then M has more explanatory force than M* for P versus P', all things equal" (Craver and Kaplan 2020: 303). Finally, they revise the above criterion by arguing that contrary to what is usually taken for granted in the literature, models are not explanations. Usually, scientists use several different models in order to "grasp" a certain phenomenon. These models are sometimes taken as describing a single mechanism but more often than not they represent numerous mechanisms or aspects of a singular mechanism which sustains a certain capability. For example, consider the circadian rhythms and how many different models are used to grasp that phenomenon or capacity. Explanations of such phenomena are rarely if ever achieved by constructing a single all-encompassing or comprehensive model. They do not reject the possibility that there could be such an "über model", they do not occur in the sciences, nor do the sciences work like that (especially the life sciences). Therefore, Craver and Kaplan identify completeness as a feature of a "store of explanatory knowledge". This amounts to a set of models which individually focus only on a portion or a single aspect of a phenomenon, but taken together, they make up a store from which the explanation of a phenomenon can be constructed. The revision, then, states that the degree of completeness is a property of the store of explanatory knowledge that is used by a scientific community (or by an individual scientist). If a store of explanatory knowledge *K* contains more relevant details about the SC mechanism for P versus P' than a store of explanatory knowledge *K**, then *K* has more explanatory knowledge than *K**.

Craver and Kaplan's proposal is their most detailed attempt to answer all the criticisms against the MDB view and the criterion of completeness of mechanistic explanation which have been accumulating steadily in the literature since MDC's paper from 2000. Although it is a notable attempt, this proposal still falls short of bridging ontological mechanisms and the ontic conception of explanation with the modeling practice of sciences and cognitive and

epistemic features of understanding and explaining. Recall Craver's postulation of the relationship between ontological mechanisms and mechanistic explanations. The things that do the explaining are "objective explanations", which are real mechanisms in nature. As Craver puts it, they cannot be more or less complete, "they just are" (Craver 2007a: 27). And real mechanisms produce real phenomena. Craver calls their descriptions explanatory texts. Hence, explanatory texts can be understood either as particular models or, perhaps better, the store of explanatory knowledge. But then, delineating the phenomenon to be explained by means of the contrastive method forces Craver and Kaplan to take a rather difficult ontological position.

Kohár and Krickel (2021) nicely formulate three arguments to show how Craver and Kaplan's proposal fails because of its background ontological assumptions: they call them *Odd Ontology*, *Multiplication of Mechanisms* and *Ontic Completeness*. Let me briefly describe them. The *Odd Ontology* should be obvious. If objective explanations are real mechanisms in nature, then the phenomena that they are causally or constitutively responsible for are also real things. But how should we understand real things that are contrastive? It is far easier to defend and understand the position asserting that P vs. P' or P vs. P* are not real entities or real phenomena rather than the one that claims they are. Second, even if Craver and Kaplan somehow overcome this problem but still uphold contrastive ontological phenomena, then there is no limit to the number of contrasts that can be a phenomenon. Any minute difference between P and P´ makes it a different phenomenon – a phenomenon in an ontic sense. However, since they adhere to the position of "Glennan's Law", this leads them to unwelcome multiplication of both phenomena and the mechanisms causally or constitutively responsible for them. Each contrastive phenomenon has its different underlying mechanism. Such a multiplication of mechanisms is certainly not a favorable consequence for anyone who wishes to embrace and defend the ontic conception of explanation. Kohár and Krickel's third objection aims at the feature of completeness. It states that "ontic, mind-independent things on their own do not have normative or evaluative properties, they are neither good, nor bad, neither complete, nor incomplete" (Kohár and Krickel 2021: 401). As Craver himself said, they just are, regardless of being complete or incomplete. In a nutshell, I believe that when we disambiguate the three senses of the notion of mechanism, it becomes evident that Craver and Kaplan's proposal fails precisely because of the conflation of the ontological sense of mechanism with the epistemic sense of mechanism. The proposal fails simply because the criteria we are supposed to consider are neither specifically epistemic nor ontic, or in other cases they judge epistemic mechanisms by criteria which are seemingly ontic and vice versa.

Kohár and Krickel's solution to the aforementioned problems recognizes this conflation of ontology and epistemology and builds upon Craver and Kaplan's account but with attention to the aforementioned three criticisms. To start, they introduce the distinction between mechanistic description and mechanistic explanatory text. A mechanistic description they take to be a neutral description of a mechanism which is not guided by a specific interest of the scientists, scientific community, or research programme. These descriptions ought to be as complete as possible (in the "more detail the better" understanding of completeness). They are what Craver and Kaplan call "store of explanatory knowledge". A mechanistic explanatory text, on the other hand, is "an answer to a particular why-question" where such a particular question requires stating a particular contrast, such as P vs. P'. These texts can be understood as particular models or sets of models of mechanisms. Kohár and Krickel then introduce a further distinction between descriptive and explanatory completeness where explanatory completeness is achieved if the model states "all and only the explanatory relevant details for P vs P´ contained in the mechanism descriptions for P and P´" (Kohár and Krickel 2021: 406). In that way, they can state that explanatory texts themselves can be contrastive too: "If an explanatory text T contains more explanatorily relevant details for P vs. P' than T* from the mechanism descriptions for P and P', then T has more explanatory power than T* for P vs. P', all things being equal" (Kohár and Krickel 2021: 407). Kohár and Krickel recognize that the completeness criterion is about models, or explanatory texts, and not about descriptions or objective texts, that is, ontological mechanisms. Furthermore, since models or explanatory texts are always answers to particular why-questions, they also recognize that the problem of saying what makes a model good or complete mechanistic model cannot be separated from the practice, purpose, and intention of a modeler. A good model of a mechanism is always a good model for a certain aspect of a phenomenon in comparison to some competitive model of the same phenomenon or of one of its aspects.

Although their proposal improves upon the deficiencies of Craver and Kaplan's proposal and introduces the contrastive explanandum to account for contrastive phenomenon, I argue, we should still not accept it. I argue that the idea of completeness of a mechanistic model should be abandoned both as a criterion and a desirable feature of a model of a mechanism. Therefore, not only does Kohár and Krickel's proposal unnecessarily complicate the picture with further differentiation of scientific explanations and our understanding of the problem, but it also retains the problem of "completeness" as a feature of explanations. They recognize that mechanistic descriptions are not complete (and perhaps they never will be), but

they nonetheless consider that the goal of a mechanistic description ought to be its completeness, where ontic mechanisms are truthmakers for the claim of completeness. Consider their example: "ideally, the description of the mechanism responsible for a neuron's firing will mention, say, how many ions and ion-channels are involved, where they are located, what size they have, etc. for every point in time of the occurrence of the mechanism" (Kohár and Krickel 2021: 404). But it is far from clear that there really is an additional goal of science different from that of constructing models for special questions and which would correspond to the complete mechanistic description independently of any of our interests (for a different view, however, see Povich 2021). Glennan seems to make the same objection in his (2017): "[The] perfect-model model is itself a bad idealization of science. Models are always partial and incomplete, and accordingly a realistic account of explanatory norms should seek to understand how partial and incomplete models can explain, rather than treating them as imperfect approximations of an ideal model" (Glennan 2017: 83). There is no clear argument for how completeness and descriptions that are independent of interests and purposes helps us in achieving understanding and explanation of a phenomenon. What do we get if we strive to have a complete mechanistic explanation of a phenomenon and what would this amount to? What would a complete explanation of protein synthesis look like? Since there is no clear argument for the benefit of the completeness of an explanation of a phenomenon (in fact, the case seems rather opposite) the completeness criterion, I believe, is a distraction which directs the discussion to a dead end.

But what, then, should be the criteria for grading one mechanistic model as being better than another in the context of a specific question or aspect of a phenomenon? Baetu thinks that there cannot be objective criteria for determining which details matter and which do not. Hence, we have what he calls the "explanatory leakage problem" (Baetu 2015: 778).[65] He presents one way that scientists try to resolve this problem. By using computational (in silico) models of mechanisms scientists are able to confirm that a proposed model of a mechanism is capable of producing the phenomenon of interest: "if the output of the mathematical model matches the experimental measurements of the phenomenon of interest, this is taken as evidence supporting the claim that the proposed mechanism is quantitatively sufficient for generating the phenomenon" (Baetu 2015: 781). In silico methods allow us to supplement a qualitative model or description of a mechanism with a set of "quantitative sufficiency" and "parameter

[65] Although he still calls this favored criterion "the completeness of mechanistic explanation".

sufficiency" (Baetu 2015: 782). At that point, then, we have evidence that one particular model of a mechanism can produce the phenomenon of interest. But notice that often this by itself is not evidence that it indeed produces the phenomenon. Recall my discussion on the evidence of mechanisms from section 1.5.2. In medicine, especially considering mechanisms of action of various drugs, population studies are supposed to offer evidence that the proposed model of a mechanism is a good representation of a causal mechanism that actually produces the effect in some population. Furthermore, notice how Baetu's discussion also accounts only for a specific aspect of a phenomenon or answers a particular question about the behavior of mechanism we are interested in. In that way, in silico methodology is used specifically to answer particular "what if things had been different" questions. But in silico methodology and computational corroboration, then, does not offer anything resembling the completeness criterion. In fact, it offers one possible criterion of a good mechanistic model and not a complete mechanistic model – informativeness about a specific counterfactual scenario. It is worth mentioning that in silico methods are now widely used in medicine in the process of finding new uses for old drugs – drug repurposing.[66] These methods, as I have noted above, offer evidence that such a mechanism is probably capable of producing the effect. Whether it will, and to what extent, is always (or at least always should be) tested by population studies. Though Baetu's proposal lacks the formality of Craver and Kaplan's and Kohár and Krickel's proposals, it rests upon scientific practice and so offers some important insights. It points to a view that scientists cherish models that are more informative than their competing models, that is, models which can predict outputs of different inputs better or more consistently.

As I have already noted, the view I accept here is what Glennan calls a minimal form of scientific realism. I do not claim that there are no causal structures underlying observable phenomena, nor do I claim that we can never grasp these causal structures with our models of mechanisms in a way that satisfies our specific explanatory and predictive needs. The claim, rather, is that the completeness of explanation, the ontic conception of explanation, and ontic constraints do not reflect what scientists actually do.[67] When we clearly distinguish separate

[66] In the next chapter, I will claim that medical treatments can be given a mechanistic explanation of their own or can be considered as models of mechanisms on their own. I will show how in silico methods play a role in such model constructions.

[67] There are of course other positions in the literature. For example, Colombo et al. (2015) adhere to a "nonrealist" position. They argue that in discussing mechanistic explanation one does not need to presuppose scientific realism either in arguing for it or against it. Furthermore, Glennan claims

understandings of mechanisms in mechanistic philosophy, we can have a better understanding of what to expect from models of mechanisms and how we can assess whether these expectations have been met. Therefore, by distinguishing the separate understandings of mechanisms in the OM, EM, and MM theses, I argue, one is not forced to accept any ontological postulate about real mechanisms when discussing mechanistic explanations of the EM thesis. Certainly, models of mechanisms from the EM thesis refer to and in some ways are either similar to some real portion of the world or they provide resources for making claims about the real world. But this does not mean that the mechanisms and phenomena of the EM thesis are the same as those of the OM thesis. Mechanism according to the OM thesis is an ontological mechanism, Craver's objective explanations, or simply a fact of the matter, while the phenomenon is an objective state of affairs or an event. A mechanism from the EM thesis, on the other hand, is referred to by an explanans and a phenomenon is referred to by an explanandum. But an explanandum is not the phenomenon itself (the real thing out there) nor is the explanans (contrary to the ontic conception) the real mechanism itself.

There is nothing within the EM thesis that is inconsistent with the claim that there is a model of mechanism responsible for the phenomenon P vs. P'. Furthermore, not only does it allow that a model of mechanism accounts for the phenomenon P vs P', but scientists will also value more highly the model which can account for more than one contrastive phenomenon. Therefore, a model M is better than a model M' if it can explain more contrastive phenomena (P vs P', P vs P*, and so on). Colombo et al. in their (2014) seem to be arguing for the same thing. They claim that "[a] genuinely explanatory model will be one that answers several counterfactual questions as well as affords opportunities for controlling and manipulating the behaviour of the target system" (Colombo et al. 2014: 191). In the next chapter I will clarify the claim that the criteria of a good mechanistic model are to be epistemic rather than ontic, when discussing prediction in mechanistic philosophy. This is to say, a good (or a better) model of the mechanism M will be able to predict and explain more counterfactual scenarios involving interventions into a mechanism than a model M'. In the next chapter, I will discuss how this proposal fares better as a description of the practice of medical sciences concerning explanation and prediction of interventions for the means of treatment, and how it offers normative constraints on the design of interventions. As I shall discuss, in philosophy, prediction has been analyzed both as a claim about the truth of a proposition or as a type of explanation of an event

(apparently, from a personal conversation) that Bechtel now calls his epistemic conception of mechanistic explanation "west coast idealism" (2017: 220).

that has yet to happen (or has happened but we are not aware of its outcome). In the latter kind of prediction analysis, prediction explains why an event is to be expected, and therefore, we have an explanandum which certainly cannot be considered as an ontic phenomenon.

2.7. Dysfunctionality and mechanisms

Dysfunctions or malfunctions are common notions in all of the applied sciences. Practice often requires dealing with broken things, either by explaining how things become broken or dysfunctional, or how broken or dysfunctional things can be repaired. To find an example, there's no need to look further than medical science. All subfields of medical science that study biological, physical, and chemical processes in the human body (immunology, endocrinology, neurology etc.) abound with talk of functions, dysfunctions, and malfunctions. Browse through any medical journal and most likely you will find the notion of function mentioned as much as the notion of mechanism. Similarly, enter the term "dysfunction" in the PUBMED search engine and you will get 2,233,185 results while "malfunction" will give you 16,300 results. With this in mind, it is rather odd that the discussion on dysfunctions or malfunctions in philosophy is not at all comparable in quantity and quality to that on functions.

The significance of the notion of function in mechanistic philosophy, however, cannot be overestimated. Recall how Glennan's law is one of the key aspects of mechanistic philosophy: there are no mechanisms by themselves, a mechanism is always a mechanism for some kind of phenomenon, a regularity, or is something that performs a function. Nonetheless, in spite of the extreme controversy of concerning the notion in philosophy, mechanistic philosophers are usually quite liberal in referring to functions. Most of the time, when discussing the metaphysical, epistemological, or methodological aspects of mechanistic philosophy, philosophers do not give a specific account of function prior to discussing, analyzing, or proposing an account of mechanism. I would add that considering the controversy surrounding the notion of function, it does seem advisable for those philosophers to put the discussion of functions to one side in developing their accounts of mechanisms. However, this does not mean that mechanistic philosophers have not discussed functions. I have already mentioned some of the views on functions of mechanisms or functions of component parts in section 2.5.2. Considering the bulk of the mechanistic literature, then, three views on functions are prominent: Garson's different functional mechanisms (SE-mechanisms, BST-mechanisms,

CR-mechanisms), the CR account of functions by Craver (2013), and the organizational account of functions (hereafter OF) endorsed by Bich and Bechtel (2021).[68] So, what can mechanistic philosophy say about cases of broken or dysfunctional mechanisms in the light of these different senses of mechanism for a function? More importantly, what does an explanation of diseases qua dysfunctional mechanisms look like?

One possible way of proceeding is to argue for a certain account of functions, figure out how biological mechanisms fit into that account, and then claim that a dysfunction is a loss of function or at least a change in a mechanism's function which leads to negative patient-related outcomes. Then, finding sufficient and/or necessary conditions for a mechanism to maintain or lose its function would be the main issue. Following such a rationale, we should expect that diseases, or perhaps better, pathological states which are identified with having a certain disease are states where "normal" biological mechanisms have stopped working, or at least have stopped working properly. Diseases, then, are nothing over and above dysfunctional or malfunctional biological (and perhaps psychological too) SE-, BST-, CR-, or OF-mechanisms. In this way, to talk about diseases is to talk about how and where normally functional mechanisms go wrong. The explanation of a broken or dysfunctional mechanism should be, then, considered against the background of an explanation of the normal function or proper working of that mechanism.

Most philosophers would accept this view on dysfunctional mechanisms, and Thagard (2003, 2006), Moghaddam-Taaheri (2011), and Garson (2013, 2018) are explicit about it. Although such an approach faces a few obstacles, I do not think there should be much controversy with what has been said so far. For example, many biological mechanisms become broken, dysfunctional, or they completely disintegrate, yet we do not consider these processes as diseases. The most obvious example is apoptosis – programmed cell death. But usually, the function losses that we do identify as diseases are accompanied with the consequential loss of one or more capacity which we find desirable. In that regard, it has often been noted that diseases, regardless of a particular definition or theory, impair some capacity that we possess as members of a certain species, and of a certain reference class such as sex or age (see, for example, Werkhoven (2019) for a dispositional account of health and disease). Hence, medical

[68] In a nutshell, the OF account says that X has a function if it contributes to the maintenance of the organization of a system, and, hence, to its own persistence. See Mossio et al. 2009, Moreno and Mossio 2015 for the OF account of functions.

treatments and means of prevention require the understanding of disease progress against the background of the proper or normal functioning of physiological mechanisms which sustain some physical or mental capacities. For example, having type 2 diabetes impairs the ability to efficiently control the levels of blood sugar. Having an erectile dysfunction impairs the ability to have sexual intercourse (and possibly leads to different psychological afflictions). Having depression impairs several abilities necessary for everyday life.[69] Understanding and explaining what "went wrong" in a specific case or in a type of cases of a disease which impairs one or a few of our capacities implies that we have an understanding of what it means to have and exercise that capacity and (usually but not necessarily) how it is physiologically grounded.

The activities of medical sciences and practice, in the end, aim to improve the health status of individuals and populations. Most likely, this goal is best achieved by treatment, cure, and care, or by prevention. Hence, by the same rationale, to cure a disease, at least in some cases, amounts to restoring the function of a mechanism which has stopped working properly.[70] To prevent a disease, on the other hand, should amount to an inhibition of a process by which a functional mechanism can or does become broken or dysfunctional. Indeed, this seems to be a rather plausible account of diseases qua dysfunctional mechanisms, and it does seem to reflect medical science and practice. For example, erectile dysfunction results from changes in functional roles of different mechanisms controlling the erectile capacity, sometimes due to dysfunctionality of the smooth muscle cell relaxation mechanism. Sildenafil citrate then treats this dysfunction by restoring the normal function by "helping" the NO-cGMP causal pathway to achieve the desired outcome through inhibition of cGMP degradation by the PDE5 enzyme. The widely used vaccine for tuberculosis Bacille Calmette-Guérin (BCG) was an efficient strategy of prevention of acute tuberculosis which, if untreated, leads to serious dysfunctionalities of lungs (but sometimes of the kidneys and brain too). Its administration contributed to developing a sufficient immune response to possible future infections – a prevention.

The issue about diseases as dysfunctional mechanisms arises only when theoretical or philosophical considerations are involved in understanding what the proper or normal functioning of a mechanism means and, as a consequence, what mechanistic explanations of

[69] Werkhoven, in his (2019) presents an example of an Olympic athlete with severe depression which incapacitates her to even get out of bed let alone perform some high cognitive or physical tasks.

[70] Sometimes, however, restoring the normal function of a dysfunctional body part will not be available or possible. For a particularly obvious example, consider amputations due to gangrenous necrosis.

diseases look like. This trouble seems to affect CR-mechanisms in particular. Recall the discussion from section 2.5.2. There, I presented Garson's distinction between mechanisms for a function and Glennan's type of minimal mechanisms, where Garson takes functional mechanisms to be a subset of minimal mechanisms – those minimal mechanisms that have a function.[71] Minimal mechanisms in general, may not have a function (there is no function of the solar system mechanism). Functions of component parts in the overall working of a minimal mechanism, if needed, should be interpreted as CR functions. In fact, Glennan's and Illari and Williamson's definitions of mechanisms are intentionally simplified and minimized in order to cover a wide range of phenomena (from geological sciences to, possibly, evolutionary biology) which can then be given a mechanistic explanation, and where talk about any non-CR functions would simply be wrong. Although Glennan and Illari and Williamson have a strongly metaphysical understanding of mechanisms, this has important epistemological and methodological implications. Even if we fail in the end, approaching to various different phenomena should be at least possible via a distinctive mechanistic epistemological and methodological framework. In fact, Garson acknowledges this: "Nothing prevents us from giving a CR-functional analysis of, say, El Niño phenomena, or demand-pull inflation, or even the way that atoms aggregate into molecules" (Garson 2018: 110). In these cases, the CR function is attributed conditionally upon the type or aspect of the phenomenon of interest. Hence, a volcano is understood as a mechanism "for" spewing lava, and if any of its parts or the volcano itself has a function, then that function is to be interpreted along the CR account of functions.

On the other hand, Garson claims that we should not view biological functions as being instantiated or supported by CR-functional mechanisms. The aspect of Garson's view I find worrisome is that he does not have a philosophical argument against CR-functions and CR-mechanisms in biology and biomedicine. He does not expose a potential flaw in the account which makes it either inconsistent or incoherent. Garson's view on the matter is perhaps best described as being motivated by his view of biological and biomedical sciences and practice. Garson argues that biology and biomedicine consider diseases as broken or dysfunctional mechanisms for a physiological (or psychological) function (either as SE-mechanisms, BST-mechanisms, or OF-mechanisms) not because of some clearly established medical theory with

[71] Let us recall what is Glennan's minimal mechanism: "A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon" (Glennan 2017: 17).

a firm philosophical basis, but rather because this is a useful heuristically in biological and medical discoveries and a useful approach to intervention for medical purposes. Here is an example: "First, when I describe anencephaly as the result of a breakdown in a mechanism for neurulation, rather than as having its own mechanism, I convey critical information about its etiology. I am guiding you to the root problem, as it were, underlying anencephaly. Second, I set up a heuristic for future biomedical discoveries. Anencephaly results from disrupting neural tube folding at the anterior neuropore" (Garson 2018: 113). Considering the CR account of functions, then, Garson identifies at least two reasons not to favor such an account in biology and biomedicine: first, it seems to be too permissible to what can count as a mechanism for a function in biological systems, and second, quite possibly, minimal mechanisms and CR-mechanisms do not allow broken mechanisms, only mechanisms causally responsible for different phenomena. But, as said, these problems do not stem from an internal inconsistency or incoherence in the CR account.

Recall that the normal functioning of a mechanism or a part of a mechanism in the CR account is not defined by a universal or statistical contribution to some behavior of a system which it is a part of, or by the effect which it has been selected for. Rather, the normal function of a mechanism or part of a mechanism is given by a scientist or a scientific community relative to a specific end point of inquiry or a phenomenon they are interested in. The functional role of a part in a mechanism is therefore relative to an outcome, phenomenon, or specific aspect of a phenomenon. This makes the CR account sympathetic to mechanistic descriptions or models of mechanisms for pathologies and this, according to Garson, should be unwelcomed for both of the reasons mentioned above. Indeed, Craver in his (2013) acknowledges that the CR account permits the construction of mechanisms for pathologies and praises it: "One can describe the function of items in the mechanisms for anoxic cell death, the production of cancer, and the progression of Alzheimer's disease" (Craver 2013: 149).[72] And, also: "When one describes an oncogene as an oncogene, one is describing it functionally without being committed to the idea that the oncogene survived by virtue of being an oncogene. Indeed, it would seem likely that it survived in spite of the fact that it functions as an oncogene" (ibid.). But notice that constructing a model of a specific disease mechanism (or a mechanism for or

[72] Darden et al. (2018) do not even question whether diseases can or should be represented as (pathological) mechanisms but apply their formal framework to mechanisms underlying genetic diseases. Similarly, Thagard in his (2005), although stating that diseases are always explained in the context of proper/improper distinction, acknowledges mechanistic explanations of diseases.

of a pathology) always includes parts which are also parts of some normally functioning physiological mechanism (and these mechanisms are thought to be individuated by the function they perform in order to sustain homeostasis in the human body).

So, with this in mind, here is the problem in Garson's (2013). An alien race decides to destroy humanity by toxic gas and proceeds to develop a model of the mechanism by which they will kill human beings. What follows then is that on "the CR theory, they would be correct, relative to those goals and interests, to say that the function of human lungs is to deposit a toxic gas into the bloodstream" (Garson 2013: 331). This does not just multiply possible descriptions of mechanisms but also has consequences which, for Garson, are rather odd: "That strikes me as counterintuitive. It seems to me that the function of the lungs is to distribute oxygen and to remove waste and that this proposition is not falsified just because someone has other plans for it" (Garson 2013: 331). Under the SE or BST account of functions lungs have not been selected in the past because they deposit toxic gas into the bloodstream, nor because they, statistically, impair one's survival and reproduction. Rather, the case is exactly the opposite. Furthermore, Garson argues that if scientists indeed implicitly accept the CR account of functions, we should expect that different sciences will have different function ascriptions to the same biological parts, depending on the interests and perspectives of those particular sciences. But, as Garson claims, we do not see this. We see the same function ascriptions in, for example, evolutionary biology and pest toxicology. Therefore, all things considered, Garson concludes that there must be a parasitic relationship of dysfunctions upon functions in the biological and biomedical sciences. Functionality, then, has at least a theoretical primacy, and dysfunctionality is only to be understood as a description of a deviation from "normal" (where normal is limited to some non-CR account of functions). But Garson's view is not without problems either. Indeed, that the CR account allows human lungs to have a function of depositing a toxic gas into the bloodstream seems to be odd or counterintuitive. But, again, I do not think it demonstrates that something is seriously wrong with the CR account. Furthermore, although it seems that the CR account is not supported or reflected in biomedical practice, SE and BST accounts of function are also not excluded for being at odds with biomedical practice in some cases. Both accounts fail to recognize some diseases as being diseases or will have trouble incorporating dysfunctions into their framework (see, for example, Davis 2000, Kingma 2007, Ereshefsky 2009 for interesting discussions).

In what follows, I will not presume that scientists who study diseases in humans and animals have a working definition of biological function grounded in a specific philosophical theory.[73] By observing the literature and practice, it could turn out that medicine is more inclined towards one or the other account of functions, but it would still be a highly disputable claim. Craver claims (and Garson acknowledges it) that the CR account is often an important methodological feature of life sciences. In neuroscience, Craver argues, searching for selective etiologies of functions often can and will obscure the discovery and explanation of different capacities. In addition, consider a view on diseases and functions in the pathology textbook already quoted in Chapter II: "'Disease is a consequence of a failure of homeostasis', where homeostasis is the concept of equilibrium within the body despite changes in the internal or external environment. If you accept this definition then understanding the mechanisms of disease will involve understanding the processes for maintaining homeostasis, identifying the agents and events that disrupt homeostasis, trying to determine why homeostatic mechanisms fail, and whether any intervention can prevent or correct this sequence that results in a disease. Rather arbitrarily, we shall decide that a disease should have the potential to produce some impairment of function…" (Lakhani et al. 2009: 3). Should this be interpreted as an SE, BST, OF, or CR account of functions? We should not, I take it, be surprised that different philosophical accounts of functions will emerge when considering the literature of biological and biomedical science and medical practice. Yet, scientists study biological, chemical, and

[73] In this dissertation, I have skipped going deep into debates on various notions and concepts and accepted some views and stances without discussion. The rationale behind this is to keep certain aspects of the discussion on mechanisms in the focus of the dissertation and its topic. I intend to do the same thing with the explication of the notion of biological functions. Even scratching the surface of the discussion on biological functions amounts to falling into a dark pit of decades-long and controversial debate in philosophy. I have no hope that I could philosophically contribute to this discussion nor that this dissertation is the right place to do this. I take that in the context of this dissertation, going into such a discussion would be highly unadvisable, undesirable, and time and space consuming. On the other hand, presenting the philosophical debate on functions without arguing for a specific theory or an account would be equally pointless. A short one would not do it justice, while a long one would be too exhaustive. Nonetheless, considering the EM thesis of mechanistic philosophy, the CR theory of functions seems to be the most suitable theory. The EM thesis is a thesis about explanatory practices, media, and concepts, and it seems to me that the CR theory offers a way to talk about functions of mechanisms and their parts without troublesome metaphysical presumptions. Whether this is so for ontological mechanisms too, I will leave for some future discussions.

physical dysfunctionalities occurring in the human (and animal) body framed in the methodologies of the science of pathology and pathophysiology.[74]

2.8. Mechanistic explanations of diseases

So how should we understand a disease qua dysfunction within the mechanistic framework?

One way to construct a mechanistic explanation of a disease is to offer a contrastive description of all or some of the steps in a model of a fully functional, or normally working physiological mechanism (regardless of the theory of function one prefers), with the contrasting case being the specific malfunction one is interested in explaining. The idea is represented in Figure 9, where a model of a mechanism is represented as a simple directed graph.

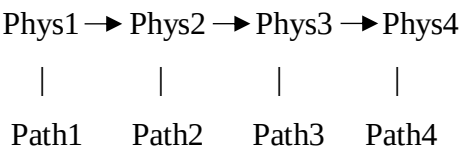


Figure 9. The stages of a physiological mechanism with their pathological counterparts. Arrows represent causal relations. Appropriated from Nervi (2010).

Let us consider MDC's definition of a mechanism as an example.[75] A pathological description, then, refers only to a certain step in the sequence from set-up to termination conditions where the usual production of regular changes has somehow been severed. A

[74] Though Boorse has claimed that his BST theory of health and disease is not only the fittest theory for biological and biomedical sciences, it is also the one that biomedical scientists (pathologists) implicitly use: "I am content for the BST to live or die by the considered usage of pathologists" (Boorse 1997: 53). Nonetheless, it is at least questionable that pathologists have a working theory of health and disease, and a corresponding one of functionality and dysfunctionality. Hesslow (1993) and Stempsey (2000) have argued not only that pathologists do not have an account of each of these notions but that they do not even need one to do what they do.

[75] "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (MDC 2000: 3).

description of a disease explains the failure of one stage to produce another stage. Figure 9 demonstrates this view. Every step in a description of a physiological mechanism has its pathological counterpart. So, for example, the "Path2" stage describes which factors of the "Phys2" stage have been changed so that the causal path from "Phys2" to "Phys3" is interrupted. In his (2010), Mauro Nervi summarizes the idea as follows: "A model of mechanism is built according to the researcher's pragmatic purpose, and the disease is explained as background of the central normal situation, which dominates the theoretical scenario. When explaining basic pathological phenomena, one needs to stress the similarity of structure between the normal mechanism and the pathological phenomena being studied. In such a case the researcher with this explanatory purpose postulates a mechanism to explain the disease: in other words, a malfunction or defect in each step of the mechanism can explain a particular form of disease" (Nervi 2010: 218). Thagard, similarly to Nervi, says that explanations of diseases "presuppose a background of normal biological operation that has broken down", and therefore, a "medical theory is a representation, possibly distributed among human minds and computer databases, of mechanisms whose proper and improper functioning generate the states and symptoms of a disease" (Thagard 2005: 59, 60). Accordingly, then, a model of a disease amounts to a contrastive explanation of normally functioning stages in a mechanism, from "Phys1" to "Phys4", with a representation of the difference between "Path2" and "Phys2".

How should this look in practice? Fortunately, the philosophical literature offers some insights. For example, in his (1993), Peter Lipton discusses four different types of factors in causal explanations: those that make a difference to malfunction but not to normal function, those that make a difference to both malfunction and function, those that make a difference to function but not to malfunction and those that do not make a difference to function or malfunction. In their (2016), Van Eck and Gervais apply this to mechanistic explanations and claim that available accounts of mechanistic explanations have not been clear on the differences between parts of mechanisms: i.e., those that are relevant to (i) malfunction but not to function, (ii) those that are relevant to both function and malfunction, and (iii) those that are relevant for function but not to malfunction. This is to say that when trying to understand a specific malfunction, some features of a properly or normally functioning mechanism will be relevant while others will not be. A good explanation of a malfunction, according to van Eck and Gervais, includes distinguishing those features or factors that make a difference to malfunction from those that make a difference to both malfunction and function. Why? Consider Lipton's

example quoted and discussed by Van Eck and Gervais: "Suppose that my car is belching thick, black smoke. Wishing to correct the situation, I naturally ask why it is happening. […] The problem is that many of the causes of the smoke are also causes of the car's normal operation. Were I to eliminate one of these, I might only succeed in making the engine inoperable" (Lipton 1993: 53). Obviously, features that make a difference to malfunction and those that make a difference to both malfunction and function are important in understanding why the system is not working properly. However, it is important to differentiate between the two because if we eliminate a feature that makes a difference to both, then we will not be able to restore the proper function. At least this is what the rationale behind van Eck and Gervais's use of Lipton's example amounts to.

Let us consider, as van Eck and Gervais do, that the specific problem in the case of a car belching black smoke is oil leakage onto a hot exhaust pipe. The exhaust pipe is a difference-making factor to both function and malfunction since it is involved in the normal functioning of the car as well as this malfunction. However, replacing the exhaust pipe will not be sufficient, that is, interventions into a factor relevant for both malfunction and function will not repair the malfunction and restore the function. Replacing the exhaust pipe will not make a difference to the occurrence of black smoke. Fixing a rupture in the oil reservoir, however, will repair the malfunction since the rupture is a difference-maker factor only for the malfunction. Similarly, no matter whether a person can achieve erection or not, the outcome will depend on certain entities and their features – most importantly, the ones in the NO-cGMP causal pathway. In order to understand how erectile dysfunction comes about, some features of entities in this pathway will matter while others will not. Usually, everything further down from the degradation of cGMP by PDE5 will not be as important to an explanation of dysfunctions of the NO-cGMP causal pathway as it will be in understanding the mechanism of vasodilation. But in order to understand how the malfunction is caused and how it can be fixed, van Eck and Gervais claim that "you need to understand the normal function of this component as well as the normal functioning of the mechanism in which the component is situated" (van Eck and Gervais 2016: 129). Only by knowing the parts and activities involved in the black smoke phenomenon can you understand where to look for the malfunction difference-maker and how to fix it. Only when the entities, their properties, and their activities constituting the NO-cGMP causal pathway had been discovered and understood could scientists understand the causal structures behind erectile dysfunction. Diseases, then, are always understood against a

background of a physiological mechanism that we label as broken, dysfunctional or malfunctional.

This rationale sometimes perfectly translates to medical science and practice, but it is doubtful that it always does so. Perhaps the failing NO-cGMP causal pathways is a nice translation of the example of an oil reservoir in medical practice. Similarly, a rupture in the anterior crucial ligament (ACL) is another example that fits into this picture. Surgery aims only at the factor responsible for the ligament dysfunction – the rupture, and the consequent treatment process will be particularly defined by the type of rupture, and so by the mechanism of injury. However, many medical interventions aim at factors that make a difference to both function and malfunction – a vivid example is an amputation of a gangrenous limb or mastectomy. In those cases, either we cannot simply identify a difference maker to malfunction and eliminate it, or the causal effect of malfunction is irreversible. Overcoming or eliminating an effect of a difference maker in such cases does not seem like a viable strategy. In the case of type 2 diabetes, an insulin pump overcomes the effect of a failed or dysfunctional mechanism, but only by acting on a factor that is relevant to function and dysfunction – blood sugar levels. A more notorious example would be a virus infection. Treatments of patients with collapsing lungs due to SARS-COV-2 infection target factors that are difference-makers to both malfunction and function – for example, pumping air by ventilators into a patient's collapsing lungs.

Nonetheless, van Eck and Gervais's argument translates to medicine up to a point. Malfunction or dysfunction in biomedicine is understood as a defect or a change in some component part or activity of a physiological mechanism, where this leads to a different or changed function of that component part or mechanism as a whole. Medical vocabulary is abundant with terms implying dysfunctions or malfunctions. That is, medical explanations of diseases are full of terms such as interruption, blockage, inhibition, drainage, deterioration and so on. Furthermore, van Eck and Gervais's characterization of factors in mechanisms also seems to be translatable to medical practice. Medical attention to dysfunctional mechanisms seeks to identify three kinds of difference-making factors: factors that make a difference to function, malfunction, and to both function and malfunction. Even if interventions are not always performed only on difference-making factors for malfunction, understanding how there are three kinds of factors in any malfunctioning mechanism guides medical intervention

strategies. Understanding which aspects of a phenomenon are related to which difference-makers seems to be a reasonable strategy for developing interventions into mechanisms.

However, it is no more than a theoretical presumption that there can always be an easily describable pathological counterpart to every stage in physiological mechanism or that changed properties of one or more component parts of a mechanism will suffice to explain a malfunction. For example, it is a presumption that the "Path3" stage in a physiological mechanism will always be a pathological counterpart description of "Phys3" (Figure 9). Consider cases where it is more convenient to understand the "Path3" stage as the outcome of the "Path2" stage, rather than the contrastive description of "Phys3". In these cases, the "Path2" stage includes parts which cannot be found in the "original" "Phys2" and "Phys3" stages of the physiological mechanism. Rather, they are products of previous pathological stages. Although Van Eck and Gervais's account is perhaps more common in medicine, cases where pathological descriptive counterparts of physiological mechanisms are not adequate for explanation are actually not rare. In fact, they happen quite often in cases of infectious diseases. I take it that the explanation of the occurrence of tubercles in lungs in acute tuberculosis is hardly imaginable as a descriptive counterpart of a proper, functional mechanism. The occurrence of tubercles are causal products of previous pathological stages going back to the set-up stage that includes a pathogenic agent – *Mycobacterium tuberculosis*. In his (2005), Thagard offers an analysis of disease mechanisms that, in a way, presupposes this: "Viral replication in itself does not produce disease symptoms, which can arise from two sorts of mechanisms. First, viral release may directly cause cell damage or death, as when the SARS virus infects epithelial cells in the lower respiratory tract. Second, the presence of the virus will prompt an autoimmune response in which the body attempts to defend itself against the invading virus; this response can induce symptoms such as fever that serves to slow down the virus replication" (Thagard 2005: 56). Considering examples of infectious diseases, then, it is not at all clear that presuming a broken-normal or dysfunction-functional view (where the pathological stage is an explanatory counterpart of a physiological stage) is more convenient and practically useful than constructing models of mechanisms of diseases as models of mechanisms in their own right.

Another issue of representing diseases qua dysfunctional mechanisms, as is depicted in Figure 9, is that diseases do not in each and every case conform to the boundaries of physiological mechanisms which have stopped working or stopped working properly. That is, a certain disease does not always affect just one particular mechanism so that every contrastive

pathological representation of each stage in that mechanisms corresponds to a full explanation of the occurrence of all symptoms and signs. Therefore, a different approach is sometimes needed, one that accepts the EM and MM theses as its starting point. Such an approach takes symptoms and signs of a disease as a phenomenon or an outcome of a certain causal structure itself, and then offers their mechanistic decomposition and localization. This is also a starting point in Nervi's argument to view mechanisms of disease as independent explanatory entities.[76] As noted, some diseases affect or correspond to changed properties of entities which lead to changed outcomes in their mutual causal relations but affected entities and their causal relations will not always correspond to a single physiological mechanism. Diseases often spread and affect different systems and mechanisms within those systems (a single disease can affect, for example, the nervous system and cardiovascular system). To illustrate this, consider Nervi's example of diabetes mellitus where "hyperglycemia [high blood glucose] following the onset of diabetes mellitus interferes with the functioning of distant structures like kidney, retina and peripheral vessels, causing the new onset of pathological chains" (Nervi 2010: 220). In such cases, the causal relations from one pathogenic state to another pathogenic state affect different physiological systems and the explanation of a pathology as a descriptive counterpart of a single stage in a physiological mechanism seems either difficult or impractical.

What these examples show is that diseases can be studied and are studied as phenomena in their own right. Scientists study how diseases are caused, how they are constituted, and equally important, how they progress. Simon, for example, says that by studying these three aspects of diseases, medical sciences offer disease models. Modeling diseases then offers further epistemological and methodological benefits for the biomedical sciences: "The behavior of this model can then be predicted based on what we know of human physiology. To the extent that a given model, by embedding the relevant causal structures, allows us to predict and affect the clinical course of a group of patients, that model will represent a (constructively) real disease" (Simon 2008: 363). We do not need to look for complicated cases to find examples of such a practice. For example, the mechanism of injury is a familiar term in orthopedics and

[76] In his paper Nervi, however, soon abandons this epistemological argument for mechanisms of diseases and ventures into a metaphysical and ontological argument. In this ontological argument he argues for the ontological independence of mechanisms of diseases, based on MDC's characterization of mechanisms, by proposing three features of mechanisms of disease which physiological mechanisms supposedly do not possess: outcome variability, no range constraint, and ambivalence. I will not criticize Nervi's ontological argument here. Moghaddam-Taaheri has offered a persuasive critique of it in her (2011), which, all things considered, I take to be correct.

sports medicine. In orthopedics, the term refers to the specific way the trauma has been caused. Zernicke and Whiting define mechanism of injury in a straightforward sense which corresponds to the definition of minimal mechanisms: "the fundamental physical process responsible for a given action, reaction or result" (Zernicke and Whiting 2000: 514). Mechanically caused injuries, such as clavicle fractures and ACL injuries, are then distinguished by the type of process that leads to an injury and the specific position of an injury. Equally important, the exact mechanism of injury often indicates a specific pathological state which on the other hand influences and points to specific interventions, that is, treatment procedures and, in the cases of such mechanical injuries, specific subsequent therapy programs.

As already mentioned, in constructing a model of a disease mechanism (or a mechanism of injury), medical scientists and practitioners follow the EM and MM theses of mechanistic philosophy. Therefore, a symptom or sign is taken to be an outcome or phenomenon of interest of a particular mechanism. Hence, a model of causal structure responsible for a symptom or sign is constructed by using the usual mechanistic methodological heuristics – decomposition and localization being by far the most important ones. Naturally, such a phenomenon will include entities which are usually component parts of physiological mechanisms, but it will sometimes include entities which are not constitutive of proper physiological mechanisms (pathogenic agents in infectious diseases, blood clots in cases of embolism etc.). A model of a disease mechanism, then, conforms to the desiderata of the EM thesis, and the methodology of constructing a model conforms to the desiderata of the MM thesis. Following the characterization of a phenomenon, functional and structural decompositions of a phenomenon ensue. At this stage, lower-level causal operations needed to produce that phenomenon are stipulated. Next, entities which could potentially be component parts are identified. In the next step, these causal roles are assigned to entities, making them component parts of the mechanistic model, and those entities that do not actively contribute are left out of the model. Nothing within the EM and MM theses (nor even the OM thesis) prevents the construction of a model of mechanism of or for a disease. Whether diseases are real or not is a different question, but no one can, I firmly believe, argue that, however we define diseases, a great deal (or perhaps all) of them result from specific causal processes or structures in the body. The problem, as I have noted, is in the notion of function, not in the notion of a mechanism. In the next chapter, I will show how medical interventions aim at either inhibiting or inducing a certain outcome, where this outcome can be regarded, depending on the stance taken by the investigators, as an outcome of a physiological, pathogenic, or pathological mechanism.

3. MECHANISMS IN MEDICINE: PREDICTION

Abstract

In this chapter I discuss prediction in medicine within the mechanistic framework – mechanistic reasoning or pathophysiological rationale. The chapter is structured as follows. First, I distinguish between prediction claims (hypotheses about yet unobserved events) and prediction activities (inferences with prediction claims as their outcomes) and apply these notions to medical science and practice. Next, I present discussions on prediction from "The New Mechanistic Philosophy" and on mechanistic reasoning from philosophy of medicine. After arguing that the proposed accounts of mechanistic reasoning from the literature do not cover all of the examples from medical science and practice, I develop my account of mechanistic reasoning. In the last two sections of the chapter, I discuss how my account applies to two instances of mechanistic reasoning in medicine: claims about outcomes of medical interventions and claims of diagnosis.

3.1. Introduction: what is prediction?

Starting with Hempel and Oppenheim's influential paper from 1948, the notion of scientific explanation has been among the most discussed topics in contemporary philosophy of science. At least since the publishing of the aforementioned paper, philosophers have argued that the primary goal of science is to construct explanations in order to achieve understanding of worldly phenomena. Hempel and Oppenheim write at the beginning of their paper that "[t]o explain the phenomena in the world of our experience, to answer the question "why?" rather than only the question "what?", is one of the foremost objectives of all rational inquiry; and especially, scientific research in its various branches strives to go beyond a mere description of its subject matter by providing an explanation of the phenomena it investigates" (Hempel and Oppenheim 1948: 135). Although Hempel and Oppenheim were certainly not the first to discuss laws of nature, causation, causal inference, and explanatory strategies, their account of scientific explanation introduced these notions and concepts into the discussion on scientific explanation in philosophy of science.

However, we look for explanations of phenomena not only to satisfy our intellectual or scientific curiosity. Understanding how or why phenomena come about is an important step in getting to interfere with, produce, or predict phenomena. In that regard, Hempel and Oppenheim claim that "only to the extent that we are able to explain empirical facts can we attain the major objective of scientific research, namely not merely to record the phenomena of our experience, but to learn from them, by basing upon them theoretical generalizations which enable us to anticipate new occurrences and to control, at least to some extent, the changes in our environment" (Hempel and Oppenheim 1948: 138). Salmon expresses similar views and writes that "[s]cience, the majority say, has at least two principal aims—prediction (construed broadly enough to include inference from the observed to the unobserved, regardless of temporal relations) and explanation." (Salmon 1998: 126). Salmon argues that both prediction and explanation are equally desirable and pursued aims of science. Neither is worth pursuing on its own. Predictions without explanations do not improve our "scientific understanding" of the world. The practical utility of science, on the other hand, is limited if our explanations cannot yield true predictions.

Perhaps science in general pursues both goals equally. Nevertheless, it seems that some sciences, like evolutionary biology, do not seek to understand the phenomena of their domain for the purposes of prediction (at least not as a clearly distinguishable goal in addition to

explanation). On the other hand, sciences that are firmly grounded in and concerned with practical consequences and considerations arguably value prediction more than explanation. Indeed, some philosophers have argued that the ability to predict rather than to explain is the principal characteristic of sciences in their practice and is often valued more in practical circumstances. For example, Carrier argues that "[s]cience in the context of practice quite naturally places heavy emphasis on foreseeing the outcome of endeavors to bring about certain products rather than epistemic virtues like causal explanation or theoretical unification" (Carrier 2014: 98). Depending on the context and the explanation, scientists in practice often argue that understanding sometimes puts emphasis on details and this, as Carrier continues to argue, "interferes with predictive strength" (Carrier 2014: 99). This is particularly apparent in the medical sciences. For example, Adam La Caze claims that medicine, and especially its clinical practice, is a "teleological science" (La Caze 2011: 81). It values outcomes the most. So if the outcome or goal of medicine is to cure and treat, then predicting the outcomes of medical treatments seems more important than explaining them. This stance on the utility of details that explanations sometimes require can be particularly evident in contemporary medicine. The EBM movement openly favors predictive success over explanatory success. After all, it represents a framework that seeks to provide answers to "questions *that*" rather than "questions *how*".

Explanation and prediction are closely connected notions. They both aim at understanding and influencing events in the world. Achieving scientific understanding can be considered a step in pursuing the prediction of future events or the intervention into phenomena in the world while true predictions often present a starting point in constructing explanations or testing theories and models. Although many philosophers have noted the importance of prediction as one of the principal goals of science, the discussion on prediction in the philosophical literature is nowhere near in quality or quantity compared to the one on explanation. Nevertheless, this does not mean that predictive inferences have not been discussed and developed in philosophy and science. There is an abundance of literature on predictive models in philosophy, statistics, epidemiology, or economy (for example, Spirtes, Glymour and Scheines (1993) is a particularly influential account in philosophy). But, as Broadbent shows in his (2013), prediction modeling does not amount to a philosophical analysis of the notion. Simply, there is no comprehensive philosophical account of prediction that seems to offer a definition of prediction and criteria to grade its quality in the same way as philosophers have been doing with the notion of explanation for decades.

What is prediction and what kind of prediction inferences are there? First of all, explanation and prediction, share many similarities. For example, prediction, much like "explanation" or "cause", is a notion that perhaps has multiple meanings. In the previous chapter, I noted that some authors claim that *to explain* can mean different things on different occasions (Craver 2014, Illari and Williamson 2011). Craver identifies at least four different meanings. He develops his account of the ontic conception of mechanistic explanation based on the fact that explanation, as he claims, can be a communicative act, a representational act, a cognitive act, and an objective structure (Craver 2014: 29). I have expressed my reasons for doubting that explanation is an objective structure (i.e., objective explanations in Craver's terminology, or an ontological mechanism in mine). However, the remaining three meanings on the list could turn out to be descriptive of various uses of the notion of explanation in scientific, philosophical, and everyday speech. Although it seems that prediction should be a rather straightforward notion, it can also mean different things. Usually, prediction is considered a claim about some future event, a claim about the occurrence, scope, or magnitude of an event that has yet to happen. For example, a weather forecast is a claim of prediction about a meteorological phenomenon in the near future, and as such, it is concerned with the probability of its occurrence and its magnitude. Broadbent (2013) takes this to be a prediction in a narrow sense.

But prediction, arguably, involves a broader scope of claims, not necessarily about the events that have yet to happen. We make inferences about events that have happened but, for the moment, we are not aware of their existence, scope, or magnitude. For example, we can guess the result of a basketball game that has yet to happen, and similarly, we can guess the result of the game that has happened, but the result is unknown to us. Notwithstanding the fact that the basketball game occurred yesterday, and one team has won, my guess about the winner is still a predictive claim. The inferential activity by which I assert the claim about the game that happened in the past with the unknown outcome, I take, is not different from guessing the result of the game that has yet to happen. What differs in these cases is the temporal orientation of inference. As Douglas argues, in making these claims, hypotheses are "to be indexed to the predictor's epistemological state rather than temporal location" (Douglas 2009: 446). Prediction claims about the occurrence, magnitude, or scope of an event that already happened or claims about the evidence waiting to be discovered or observed are sometimes called *retrodictions*. The broad sense of prediction therefore includes both prediction about future and past events – prediction and retrodiction.

3.2. Prediction claims in medicine

What kind of prediction claims does medicine assert? The answer depends on a particular goal of medical science or practice that one is concerned with. A layman's perspective on medicine is that it is the business of curing people and treating diseases. In that respect, medicine is concerned with prediction claims about the outcomes of future events. It is concerned with the outcomes of medical treatments or the policies of public health, it asserts claims about prognosis such as seizure recurrence in such and such patients, or the probability of disease occurrence in a certain population such as cervical cancer in patients with HPV infection, etc. In other words, predictive claims in medicine about future events are claims about *therapy*, *prognosis*, *harm*, and *extrapolation*. Consider how these notions are usually defined in the EBM literature. Prediction claims about therapy are claims about the effect sizes of interventions for "*patient-important* outcomes (symptoms, function, morbidity, mortality, and costs)" (Guyatt et al. 2015: 31). Prognosis is concerned with "estimating a patient's future course" (ibid.) while harm asserts "the effects of potentially harmful agents (including therapies from the first type of question) on patient-important outcomes" (ibid.). Claims of extrapolation in medicine are claims that predict the effectiveness of treatments in target populations (based on the effectiveness of treatments in study populations). These notions are all forward looking. They are not about events that have happened. They say something about the events that are yet to happen.

If the science and practice of medicine is predominantly concerned with curing and treating diseases, then an important step is to assert a diagnosis.[77] Diagnosis, as some have argued, seem to be a matter of explaining symptoms and signs. However, I provide arguments against this view and argue that diagnosis, especially a diagnostic claim inferred from the knowledge of biological mechanisms, can and perhaps should be considered a retrodictive claim. Fuller and Flores also notice that diagnosis could be taken as a prediction claim – "inferring an outcome that is not definitively known (i.e., the presence of a particular disease)" (Fuller and Flores 2016: 49) – although their discussion on predictive models does not elaborate whether it is to be understood as a retrodiction or not. In diagnosis, similarly to the example of yesterday's basketball game, the event has already occurred, yet we are not aware of its

[77] However, as I show in the last section of this chapter, establishing a diagnosis is not a necessary step to begin medical treatment.

outcome: patient *X*'s disease. Whether medical diagnosis is a retrodiction, explanation, or accommodation will be discussed in the last section of this chapter.

Since medicine is equally concerned with the health of populations and individual patients, prediction claims in medicine can be either about the population or about individual patients. Andersen notices this distinction in her (2012) and offers the following view: "The first kind of prediction is statistical: given the treatment options evaluated in available studies, which treatment results in the best distribution of outcomes for the sample patient population? The second is singular: given the distribution of outcomes in the patient populations for several available treatments, what treatment is most likely to result in the best outcome for this individual patient?" (Andersen 2012: 994). The transition from claims about the population to claims about individual patients is one of the main concerns of the EBM framework. An interesting discussion about the model of such a transition is offered by Fuller and Flores in their (2015). Notwithstanding technical terms within, the model, relies on simple induction. In this case, the simple induction strategy preferred in the EBM literature states that if the study is very similar to the target, and we have no compelling reasons to think that the results from the study cannot be applied to the target, we can apply (extrapolate) the results from the study to the target and predict the outcomes. Simple induction or extrapolation is deeply problematic, but it is still the widely preferred model of prediction from the study to target populations (see Fuller 2021).

Predictive models imply that predictive claims, be it in medicine, physics, or everyday life, are usually not guesses from the top of our minds. Most of the time we use or go through some inferential procedure or method to assert a prediction claim. As we saw in the previous chapter, a methodology for constructing a particular explanation is not the same thing as an explanation of the phenomenon itself. For example, a particular scientific explanation of some phenomenon, say the model of the mechanism of DNA replication, is an altogether different thing from the methodology used to achieve this explanation. Broadbent establishes a similar distinction concerning prediction: there are *predictive claims* and there are *predictive activities* (2013: 90, 91). To illustrate, let us again use the example of a basketball game. A predictive claim is a statement about the winner of yesterday's basketball game. The kind of inference one uses to get to this claim will be a predictive activity. These can be various. We may simply take a wild guess. We may prefer one team over the other and therefore be inclined to predict that that team has won the game. This claim may turn out to be true, but the kind of activity

used to assert this claim will be an impoverished one, or simply not good. On the other hand, we can analyze both teams' strengths and weaknesses, calculate the probabilities of outcomes based on both teams' prior games and matchups, or use some other type of inference. These are all far better inferential activities than taking guesses from the top of our minds. Predictive claims are, therefore, true or false, and prediction activities can be good or bad. Although a good prediction activity does not necessarily need to produce a true predictive claim, since it is always possible that a better team loses the game, better prediction activities often yield true prediction claims.[78]

The next set of questions, then, is the following. What kind of predictive activity or activities are there, and which ones are used by medicine? What are the criteria for a good prediction activity or inference in medicine? To answer this question we cannot, I believe, avoid addressing the question of relation between explanation and prediction. Are they just two sides of the same coin or do they involve different epistemic constraints and strategies? That is, are they different and completely separable epistemic and cognitive activities yielding different types of things? Certainly, prediction and explanation are deeply and inevitably connected. But how are they connected?

3.3. Prediction activities in medicine

Do successful predictions really confirm explanations? Not always. Do good explanations yield true predictions? Not always. As far as the philosophical literature is concerned, one can find two accounts of connection or relation between (scientific) prediction and explanation.

The first view states that prediction is a claim that serves to test the correctness or aptness of a particular scientific explanation. A similar stance on the relation between

[78] As a side note, I would like to mention that in the discussion on prediction in medicine, Fuller and Flores distinguish between prediction activities and prediction inferences. They distinguish activities from inferences based on the outcome: "a definitive forecast, as their conclusion" is the outcome of activities (Fuller and Flores 2015: 50). The difference, in their view, is that an activity does not necessarily imply that it has any outcome. Nevertheless, these nuances should not concern us here since they potentially obscure the picture and the immediate philosophical merits of the prediction activity/inference distinction are dubious. Even Fuller and Flores do not pursue further possible consequences of this distinction.

prediction and explanation can be found in, for example, the "predictivism versus accommodation" debate (e.g., Lipton 1990, Barnes 2005, 2008, Worrall 2014), or in the "realism versus antirealism" debate (e.g., Putnam 1975, Psillos 2005). More importantly for our present purposes, this view of prediction can be found in much of the literature on mechanistic explanation and mechanistic methodologies and strategies in biological and biomedical sciences. The second view of prediction in the philosophical literature takes that prediction is itself, in a way, a kind of explanation. The clearest example of this view can be found in a deductive nomological or covering law account of explanation by Hempel and Oppenheim. However, Broadbent argues for a similar view in his (2013). In discussing predictions in epidemiology, his account of a good predictive activity requires a contrastive explanation that stipulates why the prediction claim *P1* is more likely to occur than prediction claims *P2*, *P3*, and so forth. In his words: "A prediction activity is good if and only if it explains why the prediction claim is true rather than alternative outcomes identified as real possibilities by best current scientific knowledge" (Broadbent 2013: 113).

Although prediction has been asserted as either one of those things, I am not aware that anyone has argued that prediction must be just one of those things but not both (and perhaps neither of those things). In fact, I will argue later in the text that prediction in mechanistic philosophy, depending on the kind of question asked, can and should be considered both a type of mechanistic explanation itself and a claim about the existence of a certain causal structure. Whatever the view one assumes, prediction and explanation are in a tight relationship. Some predictions arise from explanations, and most explanations in the sciences are constructed from true predictions. No matter whether we claim that predictions serve to test explanations and that the merit of a good explanation is its predictive power, or that prediction just is a sort of explanation (as CL and Broadbent's accounts would have it), Bluhm's question is a reasonable one: "what kind of explanation do we need to make an *accurate* prediction?" (Bluhm 2013: 425).

In the second chapter I showed how the literature in the history and philosophy of medicine distinguishes between the two approaches to causal inference and explanation: rationalistic and empiricist approaches (see section 1.4., for example in Newton 2001, Bluhm and Borgenson 2011). In a similar way, prediction activities in medicine are also rationalistic or empiricist. Fuller and Flores differentiate between these two approaches as "inferences from theory" and "induction from experience" although they acknowledge it as corresponding to the

rationalistic/empiricist distinction. A rationalistic prediction activity infers prediction claims from theoretical considerations about a disease or human physiology, whereas an empiricist activity infers prediction claims based on the previous experience of observed cases (for example, the observed outcomes of clinical trials). In its modern conception, theoretical considerations in the rationalistic approach are conceived as evidence of mechanisms – models of mechanisms. So, consider, for example, Thagard's view (already quoted in section 2.8.) as a representative view of what medical theory should amount to today: "A medical theory is a representation, possibly distributed among human minds and computer databases, of mechanisms whose proper and improper functioning generate the states and symptoms of a disease" (Thagard 2005: 59, 60).

Before explicating these two prediction activities, I should mention that there is a third view on the relation between explanation and prediction, although it is not, to the best of my knowledge, anywhere explicitly advocated (especially not in the philosophical literature). This third view also arises from the empiricist tradition in medicine and simply assumes that, if medicine is practiced as a numeric or statistical method, there is no need for positing any relation between prediction and explanation. In this way we do not offer any explanation of a predictive claim, nor do we test some particular explanatory claim, model, or theory. Should this view on prediction and explanation be attributed to the EBM framework and practice? It seems it should not. As stated, Broadbent argues that predictive claims in epidemiology (based on evidence inferred from data, which is gathered through experimental and observational evidence gathering methods) must be construed and interpreted as contrastive explanatory claims (why *P1* rather than *P2* or *P3*). In that case, even if we accept that EBM's point of view on prediction (the simple induction model) at first does not require any relation between prediction and explanation in making a predictive claim about the treatment outcome based on the experimental or observational evidence, we are always testing a possible causal explanation of a relation between a treatment and its outcome or giving a contrastive explanatory claim about the outcome of treatment.

With that out of our way, let me begin with rationalistic prediction activity. Inferring prediction claims from medical theory has been the characteristic of the rationalistic tradition of medicine. As presented at the beginning of the first chapter, there had been several theories of disease and disease causation before the rise of the germ theory of disease (e.g., humoral theory, theory of miasma, and contagionism). Methods of treatment and claims about the

efficacy of treatments were inferred directly from these theories. Consider the humoral theory of disease. This theory viewed diseases as imbalances between four humors in the human body – blood, phlegm, black and yellow bile. In this regard, therapeutic methods aimed at the restoration of natural balances between humors. The inference about restoring the balance between humors can be regarded as a deductive inference or a deductive predictive activity in medicine. For example, Fuller and Flores present the following type of deductive reasoning exhibited by rationalistic thinking stemming from the humoral theory:

The quantity/quality of a humour is imbalanced to degree *x* (disease)

Intervention: -*x*

The quantity/quality of the humour is balanced (health)

Fuller and Flores 2015: 51

In the context of the discussion on prediction activities in science in general and contemporary medicine in particular, Hempel and Oppenheim's approach represents one possible way of doing rationalistic predictive activity or *inference from theory*. Hempel and Oppenheim's account of prediction is a consequence of their account of scientific explanation. It follows the same rationale as explanation, though it has a different temporal orientation. As they write: "[a]n explanation is not fully adequate unless its explanans, if taken account of in time, could have served as a basis for predicting the phenomenon under consideration" (Hempel and Oppenheim 1948: 138). The relation between explanation and prediction in the CL account is best referred to as *the symmetry thesis*. Salmon nicely illustrates this symmetrical relation in the following quote: "Given an event that, when it occurred, might or might not have been expected, *we explain it by showing that it could have been predicted* if we had been in possession of the explanatory facts prior to the occurrence" (Salmon 1998: 52, emphasis added). If an explanation of an event tells us why the event has occurred (being a consequence of initial conditions and laws of nature), a prediction tells us that under similar circumstances, the event will occur or at least that the occurrence of the event is expected. By knowing *why*, we know *what*.

As stated, prediction and explanation here have the same form of inference: the deductive argument of the CL account. Therefore, Hempel and Oppenheim's account at the same time offers an account of both scientific explanation and scientific prediction (the deductive argument in which one of the premises is a law of general scope), and the criteria for valid scientific explanation and good quality predictive activity. Kim explains concisely how

charitable the CL account is: "An argument-type conforming to the requirements of the covering-law model can be used as an explanation, a predictive argument, or a retrodictive argument, or for still other purposes" (Kim 1964: 361). All these purposes are satisfied if one follows the deductive structure of the CL account. The difference, so to speak, is in the temporal direction of the inference – explanation explains the event that has occurred, while prediction explains the event yet to occur.[79]

However, the symmetry thesis rarely if ever holds, and this has been one of the usual critiques of the CL account in discussions that followed Hempel and Oppenheim's groundbreaking paper. Furthermore, there is a sort of consensus among contemporary philosophers of science that the indispensable part of the CL account – the laws of nature – is rarely if ever found in biology and other life sciences (at least in the form that the logical positivist thought about the laws of nature). Within humoral theory, for example, one could possibly find laws of nature that govern relations and balances between the humors, but in contemporary medicine, and especially epidemiology, there is no such thing resembling the logical positivists' account of laws of nature, nor does it seem that these sciences even search for them. Maybe we can find such a prediction activity in history of biology and medicine (or something similar), for example in the quote by Fuller and Flores, but most certainly, reasoning in epidemiological explanations and predictions eschews any reference to laws. In addition, most philosophers of biology (and especially mechanistic philosophers) would agree that laws are not used in reasoning in contemporary biology and biomedicine (even if not explicitly noted as laws of nature).[80]

Therefore, is it safe to conclude that, as a form of rationalistic approach to prediction in medicine, mechanistic reasoning is not an instance of deductive reasoning? Many if not all philosophers working in philosophy of medicine and in philosophy of science in general think so (explicitly stated in Bechtel (2011)), while Solomon (2015) implies it is also not a type of inductive inference but rather something of its own kind. However, a closer look reveals that this perhaps is not true. Underlying mechanisms can still be used as parts of premises comprising deductive inferences about the outcomes of treatments or any other predictive claim in medicine. I present and discuss this in a section on mechanistic reasoning. For now, however,

[79] In Hempel's inductive statistical model of explanation/prediction (IS), the event would be expected to a degree equal to the probability stated in a statistical law.
[80] Cf. Haufe (2013) for an interesting view on laws in biology.

it is enough to mention that rationalistic predictive activity or mechanistic reasoning does not necessarily exclude deductive reasoning.

As I discuss in the next section, one cannot find a comprehensive account of prediction in the mechanistic literature. Mechanistic philosophers have discussed prediction sporadically and superficially. When it is mentioned, however, it is introduced and described as serving the purpose of confirming hypotheses about components in the proposed model of mechanism (parts, their operations, and sometimes their organization). Prediction, then, is best viewed as a hypothesis about correctness of a model of mechanism. Mechanistic philosophers conceive explanation and prediction as corresponding each other throughout the process of discovery of a new mechanism and the construction of a model of mechanism. Models that generate true predictions are correct models of mechanism, and correct models that offer detailed knowledge of a mechanism allow one to make accurate predictions.

Interestingly, such a view on the relation between mechanisms and predictions is shared between new mechanistic philosophers and advocates of the rationalistic approach to disease causation and "scientific medicine" in the 19th century. Recall the passage by Claude Bernard from section 1.4.1. Rationalistic medicine, as Bernard argued, allows the elimination of empiricism and the numeric or statistical method. Knowing the biological causes that give rise to a phenomenon, it was presumed, allows asserting true predictions about possible interventions and controlling or influencing phenomena. Hence, the more we know about the mechanism the more we can predict its behavior. Unfortunately, history of medicine has shown that this remains a sort of rationalistic dream, and the present-day empiricist framework – Evidence-Based Medicine – took it as a starting point in its process of reshaping the inferential and evidential landscape of modern medicine.

The EBM literature is full of examples of failed mechanistic prediction activity. In the last section of Chapter I, I have presented one example of a failed mechanistic prediction claim that had beneficial consequences. Unfortunately, these are rare. Most of the time, such failures led to negative health outcomes. From a philosophical standpoint, examples of failure of mechanistic reasoning in medicine are discussed in, for example, Howick (2011), Andersen (2012), Bluhm (2013), and Fuller (2016). These authors tend to agree that mechanistic reasoning or inferring from mechanisms to outcomes is not a reliable model of predicting outcomes of medical interventions, yet they come to different conclusions about the possible

utility of mechanistic knowledge in clinical practice.[81] However, a group of philosophers, epidemiologists and public health scientists gathered around the "EBM+" project have argued for a different conclusion: mechanistic evidence is rarely if ever sufficient on its own, but it should be considered on a par with epidemiological statistical evidence when assessing claims about the efficacy and efficiency of medical treatments (e.g., in Clarke et al. 2013, Clarke et al. 2014, Parkkinen et al. 2018, Williamson 2019, Aronson et al. 2021). The rationalistic approach or inference from theory does not seem to be a reliable prediction activity for medicine, if taken on its own. The main task of this chapter, therefore, is to provide an analysis and explanation of the inadequacy of mechanistic philosophy for prediction purposes in medicine and offer potential improvements.

So what is the dominant or reliable approach to prediction in medicine then? The empiricist approach and its relying on inductive inference, the numerical method, and observations of outcomes in populations has always been part of medical epistemology. However, only with the rise of EBM has such an approach to prediction become dominant. EBM has been developed as an encompassing empiricist evidential framework for prediction and asserting causal claims in medicine, especially claims concerning the efficacy of medical treatments (claims that a treatment has a positive average clinical effect in a population), but also for other types of prediction claims in medicine (prognosis, diagnosis, harm, and extrapolation). According to the EBM paradigm, then, the best and/or good evidence for inference about prediction claims always comes from population studies. This, however, assumes that high quality claims about causal relations are always claims about population properties or relations between populations, and not about individual patients. Consequently, the conclusions of such studies are informative for the individual, particular patient only to the degree to which the individual patient is close to the population average. The applicability of population measures to individual patients must be further inferred. To see how this two-step inference is achieved, consider the model explored and presented in Fuller and Flores in their (2015) as "the Risk GP method".

We have seen why randomized controlled trials are thought to be the least biased type of study. The double blinded type of RCTs is supposed to protect the experiment against allocation problems, it should avoid confounding, and distinguish placebo effects from

[81] Although Howick acknowledges that mechanistic reasoning can sometimes be of good quality and sufficient for prediction.

intervention/treatment effects. Even better than a single randomized controlled trial are systematic reviews or meta-analyses of numerous randomized controlled trials. Although evidence gathered from RCTs is usually regarded as the best kind of evidence, evidence from observational studies are sometimes considered better if the RCT was performed poorly. Nevertheless, all EBM hierarchies of evidence consider mechanistic evidence as low-level evidence in general, no matter the type of study used to gather the evidence of mechanisms. For example, in the hierarchy of evidence from the Oxford Centre for Evidence-Based Medicine from 2009, mechanistic reasoning occupies the bottom level for the purposes of therapy, prevention, etiology, harm, prognosis, diagnosis, etc.[82]

Why does EBM treat mechanistic evidence as low-level evidence while mechanistic philosophers argue that understanding underlying mechanisms allows one to predict and control phenomena? Why do mechanistic philosophers think we should have more detailed information about a mechanism to be able to generate true predictions while most EBM guidelines do not consider mechanisms at all? This problem does not concern only practical issues in medical sciences and practice. It is pressing on philosophy of science in general and on our thinking about the relations between explanation and prediction. Andersen's claim that this should be one of the central issues in contemporary philosophy of science is rightfully compelling: "The failure of causal explanations to match up with corresponding causal interventions should thus be an issue of concern to those who are interested in mechanisms, explanation, prediction and, especially, causal methodology" (Andersen 2012: 993). Since the mechanistic account of explanation (which also involves models of prediction based on mechanisms) has become the dominant view of explanation in biological and biomedical sciences, and since EBM has become an equally dominant view on evidence in medicine, this mismatch begs a thorough response by mechanistic philosophers.

3.4. Prediction in mechanistic philosophy

Before I continue, let me briefly repeat some of the main claims from Chapter II. I presented and discussed arguments in favor of the mechanistic account of explanation as a contemporary alternative to the CL account of scientific explanation in special sciences. Rather

[82]https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009

than searching for laws of nature (which are universal and of a general scope), scientists in biological and biomedical sciences explain phenomena by investigating underlying causes that bring about phenomena or sustain regularities. To explain how some biological and biomedical phenomena arise (e.g., protein synthesis, atherosclerosis, vasodilation, alveolar edema etc.), mechanistic philosophers argue that scientists regularly and predominantly use the concept of mechanism rather than the concept of law of nature. Biological and biomedical phenomena are supposed to be caused or constituted by biological mechanisms and explained as (regular or one-off) end products or outputs of those biological mechanisms. Mechanistic philosophy, therefore, provides an account of mechanistic structures connecting cause and effect, a framework to represent these structures, and strategies of scientific inquiry about such structures (mechanisms, models of mechanisms, and mechanistic methodology).

A model of mechanism is a kind of representation of a causal structure responsible for a phenomenon (for example, 3D models, diagrams, pictures, a set of mathematical equations, a video etc.). Recall that uncompleted models of mechanisms include numerous black boxes instead of known entities. These black boxes become more transparent as we learn more about a mechanism and the entities constituting it. Details about entities, their properties, and their spatial and temporal relations vary from the specific to the abstract. Oftentimes causal relations between entities can be highly idealized (Levy and Bechtel 2013, Love and Nathan 2015). The extent to which a much a model of mechanism is abstracted and idealized usually depends on the different epistemic constraints of a specific investigation and the requirements of its users. I presented how models and representations of mechanisms, depending on their evidential support, vary from how-possibly and how-plausibly to how-actually models of mechanism. The scope of a model of mechanism, then, depends on how applicable the model is. Usually, a model of mechanism represents a recurring mechanism found in different individuals and species (e.g., the mechanism of protein synthesis is a ready-made example in mechanistic literature).

Furthermore, mechanistic philosophers claim that mechanisms are used not only ″to describe, predict, and explain phenomena", but also, that knowledge about the mechanisms underlying these phenomena is used "to design experiment, and to interpret experimental results″ (Machamer, Darden and Craver 2000: 17). Therefore, the knowledge of mechanisms should ground certain counterfactual claims about their behavior. In other words, if "one knows how a mechanism works, one can say how it would work if it were placed in different

conditions or given different inputs" (Craver and Darden 2013: 6). Or, as Woodward states: "[T]he representation [of an acceptable model of a mechanism] allows us to see how…the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves" (Woodward 2002: S375). I have claimed that this should be the main criterion of aptness of a model of mechanism. Hence, knowledge about a mechanism's output, its parts, their mutual relations, and its overall organization allows for a design and development of means of intervention to manipulate the mechanism's behavior and output. But there is a big step from knowing how a certain mechanism works to being able to predict counterfactual scenarios involving that mechanism, given different environmental settings and values of its inputs. That we should be able to achieve this by having a detailed knowledge of mechanisms asserts a metaphysical claim too: it asserts that we should be capable of taking mechanisms out of their environment and transporting them into a new one while having all of their features, operations, functions, and other characteristics retained.

Unfortunately, the discussion on mechanistic prediction in the mechanistic literature is sparse and not particularly helpful for the resolution of the issue of low reliability of mechanistic reasoning. Simply, mechanistic philosophers never really bothered with prediction. The reason, I believe, is not at all obscure or mysterious. Mechanistic accounts of causality and explanation have been on the rise in philosophy of biology and neuroscience from the 1990s. Since then, mechanistic philosophers did not discuss in detail how the knowledge of biological mechanisms is used to treat diseases. To them, this question was not of particular interest. Why? First and foremost, mechanistic philosophy in the 1990s and the 2000s started as an account of scientific explanation in the aforementioned fields of the life sciences.[83] Although Craver and Darden start their book (2013) with the example of the mechanism by which *curare* (a poison made from certain plants found in South America) kills infected victims, the paradigmatic examples of biological mechanisms in these early days of mechanistic philosophy were phenomena such as protein synthesis, cellular metabolism, and photosynthesis. In a nutshell, most mechanistic philosophers share a similar view about the inference of prediction claims based on the knowledge of mechanisms. The following claim sums this view up: having a detailed knowledge of the inner workings of a mechanism allows

[83] Although, as we have seen, Glennan's work has often been more metaphysical than that of other mechanistic philosophers. For example, in his (1996) he claims to offer an account of causation and an answer to Hume's problem of induction while in his (2017) he develops a book length account of metaphysics and epistemology of mechanistic philosophy.

one to predict, although to a degree, how that mechanism will behave when some of its constitutive parts are intervened upon, or when input values are different. Such a view is explicitly stated in Machamer, Darden and Craver (2000), Woodward (2002), and Craver and Darden (2013). In medicine, this means that inferences about prognosis, diagnosis, treatments, and their outcomes should be based on the models of biomedical mechanisms. The knowledge of causal and organizational relationships that sustain the input-output regularity in mechanisms is supposed to ground inferences about these relationships when inputs or outputs of the mechanisms are changed. Put differently, it is a type of reasoning that takes "previously established mechanisms as evidence to adduce other conclusions" (Aronson 2018: 1166). Mechanistic knowledge provides grounds for a certain kind of causal claims – prediction of a mechanism's outputs in a yet unobserved condition or circumstance. Mechanisms, therefore, offer truth conditions for making counterfactual claims about medical interventions.

There is no symmetrical relation between explanation and prediction within the mechanistic framework as this is present in the D-N account, but these two are still in a very close relationship. Mechanistic philosophers usually take that both explanation and prediction correspond to each other throughout the process of discovery of a mechanism underlying a phenomenon and the process of constructing a model of mechanism. Hypotheses about covert entities inside black boxes, their properties, and how they are spatially, temporally, and causally organized to produce a phenomenon are tested by interventions into mechanisms. Identifying an entity or entities in black boxes usually enables us to generate new hypotheses about the behavior of a mechanism's parts (and that mechanism's overall behavior as well). Quite simply, the rationale is that if the predictions of outcomes of these interventions are failing, then there is something in our model that does not correspond to or wrongly represents something in nature. Darden describes this process as involving a two-step methodology: "Reasoning in the light of failed predictions involves, first, a diagnostic process to isolate where the mechanism schema is failing, and, second, a redesign process to change one or more entities or activities or stages to improve the hypothesized schema" (Darden 2006: 30). Successful predictions increase our confidence that entities and their causal relationships are correctly represented in the proposed model of mechanism. Failures of predictions, on the other hand, suggest revisions of the model.

The claim, then, is that the more we know about entities and their mutual causal connections, the more we can predict the outcomes of the interventions we perform on them.

If NO was the entity in the black box, then certain effects of interventions would be expected. By the mid-1980s, it has been observed that NO and EDRF have similar properties. In vitro studies showed that NO stimulated the activation of sGC. These findings allowed researchers to develop new means of interventions for testing whether NO is really EDRF. In a series of unrelated experiments from different research groups, it was demonstrated that amino acid L-arginine is involved in the synthesis of both NO and EDRF, and that both manifest the same properties in reactions with different molecules. Hence, it was expected that admission of nitrates should have some positive effect in patients with angina pectoris or hypertension. And it did, but the problem for exploiting this fact for medical purposes, as I have presented in Section 1.7., was the rapid decrease in response to the admission of nitrates.

A model of mechanism is tested by intervening on the mechanism's components, by changing the relations between them, or by changing the phenomenon it brings about (e.g., these methods are described in Craver 2007a). Prediction claims about the outcomes of manipulation and control of the mechanism's parts figure as hypotheses that govern the search for the evidence of existence of these parts, their properties, and their mutual relations in a particular mechanism. In their (2013), Craver and Darden present three kinds of experiments on mechanisms. These experiments include three modes of interventions when constructing a model of a mechanism. In the first mode of intervention, we test how some component part is related to another component part, whether causally, spatially, or temporally. The second mode of intervention tests whether and to what extent some component part is relevant to overall output. So, Craver notes that "[t]hese tests involve not only revealing correlations among the states of different parts of a mechanism but, further, intervening in the mechanism and showing that one has the ability to change its behavior predictably" (Craver 2007a: 93). That is, interventions in a mechanism have the goal of expanding our knowledge of causally productive steps of a mechanism. Craver in his (2007a), for example, embraces Woodward's interventionist account of causation, both as a descriptive account of scientific method and as a normative account of causality for thinking about mechanisms. This approach requires that certain epistemic and methodological constraints be satisfied (e.g., modularity of a mechanism, or modular subassembly). Steel (2008) also argues that biological mechanisms are investigated and understood primarily as being modular. The third mode of intervention Craver and Darden identify as the research into interlevel relations within a mechanism, bottom-up and top-down relations, and their overall contribution to the output.

The relationship between the construction of a model of mechanism and prediction claims based on it is usually characterized as "an extended, iterative process" (Abrahamsen, Sheredos and Becthel 2018: 239). A model of mechanism is constantly revised depending on the observed outcomes of predictions which are in turn based on a model of mechanism. In that regard, Darden argues that discovery in biological sciences does not resemble changes or shifts of Kuhnian paradigms but rather an "error-correcting process" or "iterative refinement" (Darden 2006: 272, 306). Similarly, Craver and Darden write: "[D]iscovery of a mechanism is a piecemeal *iterative process*, not a *linear march* from constructing a schema to demonstrating its adequacy. Often anomalies turn up, and some require revision of a hypothesized mechanism schema" (Craver and Darden 2013: 201, emphases added). Therefore, the more a model of mechanism corresponds to some real thing, the more accurate our predictions become: "When a prediction made on the basis of a hypothesized mechanism fails, then one has an anomaly and a number of responses are possible" (Machamer, Darden and Craver 2000: 17). We are more certain of hypothesized entities in black boxes if predictions based on their presence turn out to be successful.

On the other hand, if a model of mechanism fails to make accurate new predictions, then something in that model does not correspond (or is not sufficiently similar, or informative) to something "out there", so we must make revisions: ″If the anomaly cannot be resolved otherwise, then the hypothesized schema may need to be revised″ (Machamer, Darden and Craver 2000: 17). Douglas asserts a similar claim in her interpretation of the relationship between mechanistic explanation and prediction: "mechanisms help provide the intelligibility that enables one to track down where an explanatory schema has failed, producing a flawed, inaccurate prediction" (Douglas 2009: 456). According to Douglas, then, mechanistic explanations are particularly useful because they generate new predictions. Darden describes four stages of anomaly resolution in biological and biomedical research: "(1) confirm the anomalous data, (2) localize the problem, (3) resolve the anomaly, and (4) assess the resulting theory" (Darden 2006: 212).

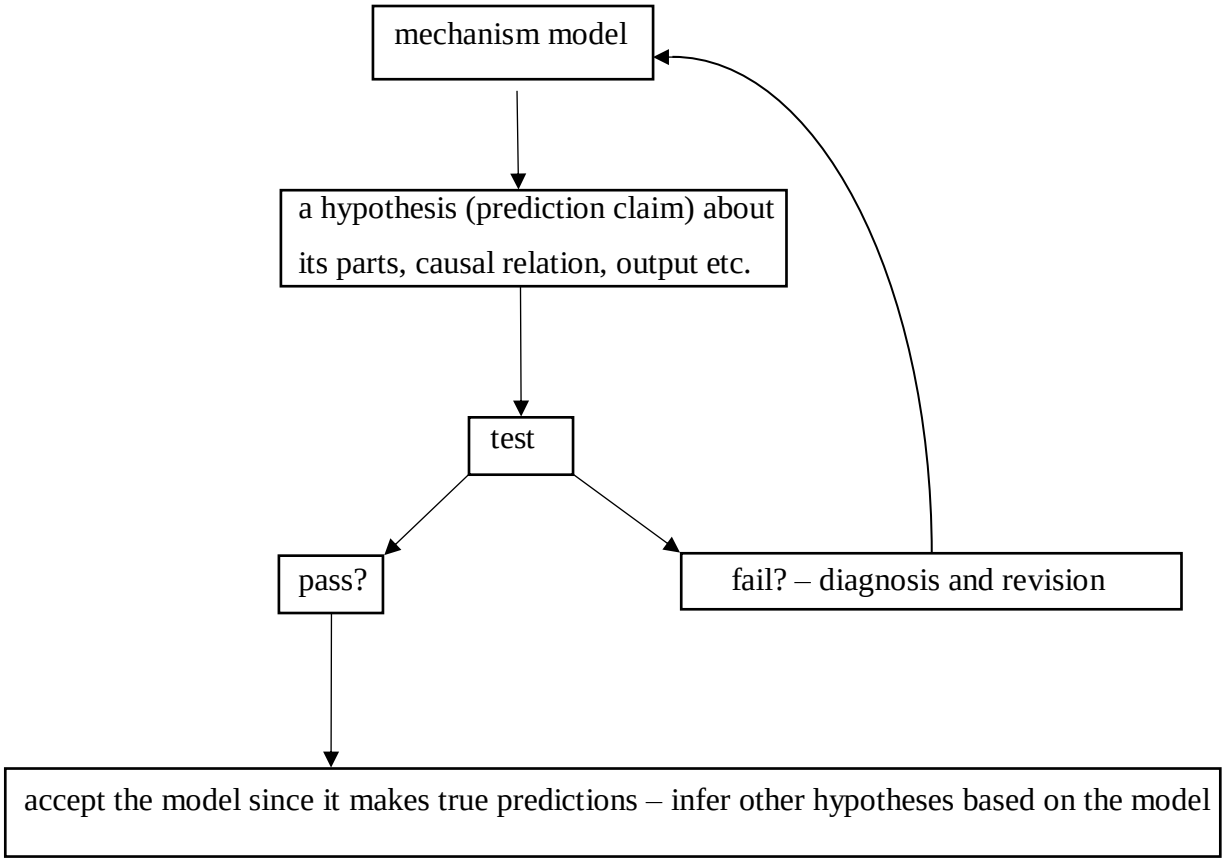


Figure 10. Prediction as a hypothesis about the correctness of a model of mechanism

It is unlikely that any mechanistic philosopher would defend the claim that a complete model of mechanism (if such a model is ever achievable) allows us to predict the behavior of a mechanism in all conceivable conditions or as a result of all possible interventions on its component parts. Surely, that would not be an easy position to defend. Although Craver has argued that the understanding of mechanisms requires specificity of details, even he takes that some mechanisms can turn out to be too complex for our models to yield successful predictions. Numerous biological mechanisms are extremely complicated and intertwined with other mechanisms. Some mechanisms, biological or otherwise, can turn out to be inherently stochastic. Craver therefore recognizes that "[s]ome mechanisms are so sensitive to undetectable variations in input or background conditions that their behavior is unpredictable in practice" (Craver 2007a: 217). Craver also recognizes that there are good and valuable

models of prediction that do not offer explanations, mechanistic or otherwise, of the phenomena they predict.

The main idea shared by all these quotes from the mechanistic literature, however, is not that having a complete explanation of a mechanism allows us to predict its behavior in all as-of-yet unobserved conditions. Rather, having a schema by which we understand how a mechanism works allows us to come up with, infer, or imagine counterfactual claims about its behavior: "to say that one stage of a mechanism is productive of another (as I suggest in Machamer et al. 2000; Craver and Darden 2001), and to say that one item (activity, entity, or property) is relevant to another, is to say, at least in part, causal relevance and manipulation that one has the ability to manipulate one item by intervening to change another" (Craver 2007a: 93, 94). That is, knowledge of mechanisms manifested in the construction of good models of mechanisms gives us grounds to assert counterfactual claims concerning interventions into mechanisms but not the extent to which these predictions can be successful.

The relation between explanation and prediction in mechanistic philosophy seems to be of the first kind (mentioned in section 3.3.): prediction claims are hypotheses that serve as tests of correctness, goodness, or aptness of a particular scientific explanation. Constructing a mechanistic schema includes testing hypotheses based on preliminary mechanistic sketches. Predicting the outcomes of interventions into mechanisms based on these sketches is a crucial part of constructing a model of mechanism. Mechanism schemas, then, allow for making new hypotheses on how to intervene into mechanisms in order to achieve the goals of prediction and control (for example, to achieve desirable patient-related outcomes). Identification of the role of the NO-cGMP causal pathway in the mechanism of smooth muscle cell relaxation provided scientists with a hypothesis about the role of dysfunction of this pathway in the development of cardiovascular diseases and furthermore, about potential targets of medical interventions into this pathway. Mechanistic explanations, therefore, are not only useful causal explanations of biological or biomedical phenomena but also *the means* of generating new hypotheses. This cannot be denied even by the most ardent proponents of empiricist medicine. After all, the majority (but still, not all) of causal hypotheses tested in population studies are based on some prior mechanistic hypothesis: either there is a mechanism or mechanisms linking the exposure and the outcome, or there cannot be a mechanism linking the exposure and the outcome.

Mechanistic literature for the most part remains silent on how the knowledge of underlying mechanisms leads to making prediction claims important for biomedical sciences (specifically the claims about treatment efficacy). The philosophy of medicine literature, on the other hand, offers some insight but with numerous ambiguities. Therefore, the inference about counterfactual claims relevant for medical purposes based on the knowledge of underlying mechanisms is the topic of the next section.

In the sections to come, I explain how the rationale described above equally applies to mechanistic reasoning and the pathophysiological rationale. Models of mechanisms serve as hypothesis-generators – predictions about the efficacy of treatments. However, there is an important difference. As we have seen, mechanistic philosophers take failed predictions as indicators of an incomplete or failed model of mechanism. But examples from the biomedical sciences show that the relation between mechanistic explanation and prediction is not so straightforward. In the biomedical sciences, failures of predictions based on mechanistic reasoning are not always considered anomalies of a model which require its revision. Predictions of outcomes of interventions into mechanisms for treatment purposes are often false and how to proceed from these failures cannot be assessed by providing simple straightforward strategies. Constructing a mechanistic prediction, as I will show, requires additional conditions that are not just hypotheses about known mechanisms. It will be claimed that mechanistic predictions of the outcomes of medical treatments require constructing novel mechanistic models – models of intervention mechanisms. This suggests that the relation between the two is more complex than previously described. It also suggests that, in the case of predicting outcomes of medical interventions, the relation between explanation and prediction should be viewed along the lines of the second view. Mechanistic prediction of the outcome of medical intervention is not just a hypothesis about a mechanistic model. It is a kind of causal explanation itself, however, with a different end-product and with some additional constraints. Fuller offers a similar interpretation when claiming that "[m]echanistic prediction is inferentially identical to mechanistic explanation; what differs are the components that are taken for granted" (Fuller 2016: 103). In his account, prediction takes a model of mechanism for granted and then reasons through it to get to the output. In explanation, the output or the phenomenon is taken for granted (although our understanding of the phenomenon changes as we find out more details about the underlying mechanism) and we try to develop a model of mechanism that could be causally responsible for it. Both explanation and prediction, then,

involve reasoning about productive steps of a mechanism and how those steps eventually lead to the mechanism's output.

3.5. Mechanistic reasoning in medicine or pathophysiological rationale

Mechanistic reasoning is used to assert predictive claims about the outcomes of medical treatments, prognosis, diagnosis, and extrapolation. However, the EBM movement considers it low-quality evidence, and puts it at the bottom level of evidence hierarchy, along with expert opinion. Predictive claims based on this kind of inferential activity as EBM proponents often claim, are unreliable and fallacious. But the EBM literature does not provide a detailed discussion on why and how such reasoning fails. The argument for the low-quality of mechanistic evidence is presented as a conclusion of enumerative induction – arguments against mechanistic evidence are based on observations of various cases where mechanistic reasoning failed to produce accurate predictions, and which, then, supports the generalization to all cases of mechanistic reasoning.

Although mechanistic reasoning or the pathophysiological rationale is an instance of the rationalistic tradition in medicine, it does not necessarily include a specific theory of disease causation (for example, the humoral theory). Rather, it is an inferential practice that takes into consideration facts about known underlying biological mechanisms, and then, reasons about their effects. In an article from 2010, Howick, Glasziou and Aronson define mechanistic reasoning as follows:

> Mechanistic reasoning is the inference from mechanisms to claims that an intervention produced a patient-relevant outcome. Such reasoning will involve an inferential chain linking the intervention (such as antiarrhythmic drugs) with the outcome (such as mortality).
>
> Howick, Glasziou and Aronson 2010: 434

The explanation of a dysfunctional (along with a functional) mechanism due to heart attack is not by itself evidence that a specific antiarrhythmic drug will have a desirable patient-relevant outcome. On the other hand, such an argument treats the reasoning based on our mechanistic model of defibrillation of a stopped heart as mechanistic evidence. In this case, a description of entities, their properties, their activities, and their interactions which together are

constitutive of the human heart enable a high-quality instance of mechanistic reasoning – that is, a piece of evidence that medical practice can use for treatment purposes.[84]

But how do these two cases of mechanistic reasoning differ? Why did the former case not produce a true prediction, whereas the latter case did? Howick et al. supplements the characterization of mechanistic reasoning stated above with the following conditions. First, our understanding of the relevant mechanisms involved and the productive stages therein that lead to the restoration of the proper functioning of the heart must be detailed and stable enough. Second, we must have enough evidence that sufficiently similar mechanisms are shared by the majority if not all members of population. Third, this mechanism must not be stochastic and unpredictable in practical considerations. When satisfied, these conditions ensure high-quality mechanistic prediction of an intervention's outcome. Of course, defibrillation has never been tested in randomized trials but, according to Howick et al., even if it were possible, there was no need for it in the end – the inference produced a stable prediction.

This understanding of mechanistic reasoning and the above stated conditions are restated again in Howick's (2011a) and (2011b). By now, the majority if not all the authors discussing mechanistic reasoning refer to this understanding, or some of its variants. For example, in a recent article where they argue for evidential pluralism in the research on treatments for COVID-19, Aronson et al. take mechanistic reasoning as one "which appeals to features of the mechanisms by which the intervention is hypothesized to lead to the outcome and to the mechanistic studies that investigate these features" (Aronson et al. 2021: 685). Similarly, Fuller refers to it as "*Reasoning through a mechanistic model*" which he defines as "a cognitive or inferential activity in which we simulate or animate the operation of a mechanism mentally" (Fuller 2016: 103). Fuller attributes such an understanding of mechanistic inference to Bechtel and Abrahamsen in their (2005) article. Indeed, these understandings of inferring outputs of a mechanism by reasoning through its productive steps reflect how mechanistic philosophers usually consider scientists' understandings of a mechanism's productive activity and reasoning through mechanistic models.

Philosophy of medicine, as I showed in the previous section, is not a particularly informative about the structure of reasoning through the model of a mechanism for asserting

[84] This motivates Miriam Solomon to argue that Howick's views on mechanistic reasoning or evidence based on the knowledge of mechanisms do not place such evidence on any specific level within the evidence hierarchy.

predictive claims. That is, there is no comprehensive account of the inference from evidence of mechanistic causal structures to claims that an intervention produces a patient-relevant outcome. Similarly, discussions presented in Howick et al. (2010) and in Howick (2011a, 2011b) do not thoroughly analyze features of such an inferential activity but they do offer the three aforementioned conditions or criteria to grade its quality. Other discussions of mechanistic reasoning or rationalistic inference in the literature on philosophy of medicine, for example in Thagard (2003), Russo and Williamson (2007), or Clarke et al. (2013), similarly eschew going into details about the steps of inferential activities included in mechanistic reasoning.

The lack of detailed discussion on constraints and features of mechanistic reasoning in medicine (in contrast to empiricist EBM's inferences) leaves Howick's account of mechanistic reasoning liable to various criticism. Out of these, the most detailed criticism is stated in Solomon (2015). There, Solomon argues that Howick's account of mechanistic reasoning offers nothing over and above the deductive reasoning that includes certain biological facts in its premises. Solomon's critique can also be applied to some other examples of mechanistic reasoning from the literature. Nevertheless, I will argue that part of Solomon's critique lies in a rather partial and perhaps not so generous reading of Howick's discussion of the subject. That is, I will show that Howick's account of mechanistic reasoning offers a way of answering Solomon's critique. However, to my knowledge, no one, including Howick himself, has made attempts at developing such a project in detail.

Let us then consider how Howick's account of mechanistic reasoning is interpreted as deductive reasoning. The first example is presented by Miriam Solomon in her (2015). She considers Howick et al.'s (2010) example of the treatment of goiters not as an example of mechanistic reasoning with some distinctive features but as a deductive inference. The argument proceeds as follows:

Premises

(1) For all x, if x is a goiter then x can be shrunk by radiotherapy

(2) Large goiters impair respiratory function

(3) Small goiters do not impair respiratory function

(4) Radiotherapy for goiters is safe

Consider a large goiter G that impairs respiratory function. Following premise (1) and premise (4) it can be safely shrunk by radiotherapy until it is a small goiter, and, following premise (3) we expect that:

Conclusion

Respiratory function will be improved if large goiters are treated by radiotherapy.

Solomon 2015: 122

Indeed, large toxic goiters have been treated with radioactive iodine since the 1940s and the 1950s, and it remains a preferred treatment of toxic goiters over surgery. The reasoning was valid, and the conclusion was true since the biological facts in the premises were correct. The radioactive iodine destroys abundant thyroid cells without presenting risk of serious adverse effects.

The second example of deductive reasoning disguised as mechanistic reasoning can be found in Thagard's discussion in his (2003). In this paper Thagard considers the role of the notion of a molecular pathway in the process of developing a mechanistic explanation of biomedical phenomena, and how it is used to reason about pathology and pathophysiology. He argues that thinking in terms of functional and dysfunctional pathways helps us in conceiving and developing treatments for diseases. His paradigmatic example of the pathway concept is glycolysis – the pathway which converts one molecule of glucose into two molecules of 3-carbon compound pyruvate and two molecules of ATP. Medicine considers many diseases to arise, according to Thagard, when such pathways become dysfunctional and where these dysfunctions are either due to molecules in pathways or to reactions between these molecules. Thagard claims that when we know which molecule or reaction is defective, we should be able to use this knowledge to "repair" dysfunctions. Here is how Thagard presents this type of reasoning:

Pathway Stimulation Treatment Strategy

Treatment question:

How can a **disease** affected by an underactive **pathway** be treated?

Treatment discovery strategy:

Determine the **molecules** and **reactions** in the **pathway**.

Identify a **molecule** in the **pathway** susceptible to increased **activity**.

Search for **drugs** that increase the activity of the **molecule**.

Thagard 2003: 246

Thagard presents his reasoning on molecular pathways as a part of constructing a mechanistic explanation and reasoning about treatment methods and their outcomes. Nonetheless, if we reformulate this in even more generic terms, Thagard's *Pathway Stimulation Treatment Strategy* does not differ substantially from deductive reasoning presented in Solomon's example of mechanistic reasoning in treating goiters with radiotherapy. So, let me reformulate his strategy:

1) Pathway *P* is dysfunctional due to decrease in activity of molecule *X*.
2) Drug *Y* increases the activity of molecule *X*.
3) Therefore, drug *Y* treats disease *D*.

Thagard's notion of a causal pathway and its use in medical science can easily be applied to reasoning behind the use of sildenafil citrate into the NO-cGMP pathway. Recall that the NO-cGMP causal pathway seems to be dysfunctional in some cardiovascular diseases or in some instances of erectile dysfunction and that this leads to the manifestation of symptoms of those diseases. As presented in Chapter II, sildenafil citrate is used to inhibit the dissolution of cGMP by binding to the PDE5 enzyme in the NO-cGMP causal pathway and thereby enhances the pathway by increasing the concentration of cGMP. The rationale was to allow smooth muscle cell relaxation and therefore enhance the blood flow in patients with angina pectoris. Translating this to the above formulation we get:

1) All angina pectoris cases result from inadequate blood flow to the heart.
2) Dysfunction of the NO-cGMP pathway impairs the widening of blood vessels.
3) NO-cGMP pathway is dysfunctional due to a decrease in the activity of cGMP.
4) Sildenafil citrate increases the activity of cGMP by binding to PDE5.
5) Therefore, sildenafil citrate allows smooth muscle cell relaxation and the widening of blood vessels and thus it is a treatment for angina pectoris.

Should we agree with Solomon, and accept that mechanistic reasoning, at least when interpreted in her terms, does not differ from deductive reasoning (although lacking laws of nature)? Quite plausibly, the examples from Solomon and Thagard suggest such a view of mechanistic reasoning. But, interestingly, all of the aforementioned authors so far imply that mechanistic reasoning is not (or at least should not be) an instance of deductive reasoning. Solomon concludes that "Howick owes us an account of mechanistic (rather than logical) reasoning, one that can justify assessments that some cases of mechanistic reasoning are

stronger than others" (Solomon 2015: 122). Although Howick et al. offer three conditions that a high-quality mechanistic reasoning should satisfy, it is still an oversimplified account that provides no persuasive answer to Solomon's question. This simple take on mechanistic reasoning, apart from acknowledging that it sometimes produces false prediction and sometimes true prediction, cannot explain why mechanistic reasoning behind the use of antiarrhythmic drugs was of a low quality, whereas reasoning behind the defibrillation of a stopped heart seems of a better quality. The latter rather successfully predicts the outcome of intervention even if it has never been tested in any sort of clinical trial, whereas the former has failed and led to deadly consequences.

However, there is a further claim implicitly stated in Solomon's quote that I will discuss next. Obviously, if there is such a thing as mechanistic reasoning in medicine, Solomon expects it to be something rather different from deductive reasoning. But it seems that she does not consider it as a type of inductive reasoning either. Rather, her argument against Howick's characterization of mechanistic reasoning is motivated by her stance that it should be something entirely of its own kind – irreducible to either deductive or inductive reasoning – and she critiques Howick for failing to deliver a satisfying account (or any account for that matter). If this is true, the question remains: what kind of reasoning is mechanistic reasoning?

In the following sections, I offer my answer to this question. I give my characterization and analysis of mechanistic reasoning, as well as the practices associated with it. This task is divided into two parts. First, I answer Solomon's question regarding what mechanistic reasoning is and what distinguishes it from the empiricist tradition, revised and expanded in the EBM framework. I have anticipated that Howick does not offer an answer to Solomon's question but that he does offer a way to start constructing one. Here, I expand this view into a more coherent argument. Answering this question will also include positioning mechanistic reasoning in the practice of medicine: I answer where we can expect to find mechanistic reasoning and what contributions we can expect from it within medical science and practice. Second, I develop my detailed characterization of mechanistic reasoning. I offer a normative account of mechanistic reasoning. That is, I discuss what should be the features of mechanistic reasoning and I lay down the constraints and conditions that it needs to satisfy. I will argue that these constraints and conditions differ depending on the types of questions from medical practice (e.g., for prediction of the outcomes of interventions or diagnosis or extrapolation of the results of experimental studies). Finally, offering a normative account of mechanistic

reasoning will then allow us to answer another question: what separates low quality from high quality prediction claims resulting from mechanistic reasoning?

3.6. Redefining mechanistic reasoning

My view of mechanistic reasoning is similar to the view of the mechanistic/difference-making difference in assessing causation from Chapter II. That is, I accept Howick's assertion that mechanistic reasoning "is best understood when contrasted with what the EBM movement believe provides the strongest evidential support, namely comparative clinical studies" as a starting point (Howick 2011a: 124). Therefore, in its broadest sense, I view mechanistic reasoning as a type of predictive inference distinguished from modern empiricist inferences in medicine by the domain or kind of facts that we reason upon, the methods we will use in that prediction activity, and finally, the particular structure such an inference will possess. Hence, one way to characterize mechanistic reasoning, and the one I will explore here, is by following Howick's assertion above: by contraposition of mechanistic reasoning to the EBM evidential framework.

The main claim of this section is the following: There is no idiosyncratic epistemology involved in predictive activity in the EBM framework that cannot be found in other sciences, and similarly, there is no distinctive mechanistic reasoning. Surely, the evidence that EBM praises the most is about the relations between exposures and outcomes on the population level, but the inferences drawn from the evidence are inductive (Bayesian approaches, statistical etc.), hypothetico-deductive approach, and even inferences to the best explanation. For example, Djulbegovic et al. claim that rather than being a philosophical (epistemological) theory, EBM is "a continuously evolving heuristic structure for optimizing clinical practice" (Djulbegovic et al. 2009: 158). Different inductive approaches, inference to the best explanation, and, as we have seen, even deductive reasoning are to be found in the mechanistic approach too. The two approaches, therefore, are not distinguished by a particular reasoning strategy, for example, deductive vs. inductive, or analogical vs. statistical. Rather, the differences between the two approaches to prediction inferences consist in:

(i) the domains of interest or reasoning,
(ii) evidence used for reasoning,
(iii) the criteria for grading the quality of reasoning.

Let me characterize this distinction in a bit more detail. What are the domains of reasoning between the two approaches? I have elaborated on this distinction in Chapter I, but I will go briefly through it again, however, this time I will focus on prediction inferences rather than explanatory inferences. Empiricist inferences that EBM holds in the highest regard concern over statistical concepts and comparative or contrastive relations between populations of patients. Mechanistic reasoning, on the other hand, takes biological, chemical, and physical processes occurring inside the body as its factual grounding for making inferences about unobserved cases. This is, of course, nothing new. Similarly, Fiorentino and Dammann propose in their (2015) that the empiricist approach of epidemiology and clinical epidemiology measures the correlations between variables and then, by using different statistical tools, proposes causal hypotheses to account for these patterns. The mechanistic stance, on the other hand, they claim, offers a biological explanation of these hypotheses. It does so by referring to its intra-individual manifestations in terms of biological causes. We can also state that the mechanistic approach goes into the individual, or as De Vreese et al. claim, it requires "parsing an individual in terms of his or her biologic make-up rather than externally observable characteristics and behaviors" (De Vreese et al 2010: 374). Instead of measuring and comparing the outcomes of interventions and the outcomes of, for example, placebo administrations in multiple cases (patients), mechanistic reasoning offers *a causal explanation* of the stages between the administration or intervention and the outcome in terms of the biological mechanisms linking the intervention or exposure and the outcome. Again, Glennan's phrase of "looking under the hood" seems like a suitable illustration of such a strategy. Biological, chemical, or physical mechanisms are what we expect to find by lifting the hood. Understanding these mechanisms is supposed to give us knowledge of how *X* causes *Y*.[85]

This leads to the second point that locates the difference between the two in the approach and understanding of the notion of the black box. In his 1998 article, Douglas Weed argues that, considering epidemiological research and its evidence, the black box should be interpreted as a metaphor for an individual. Recall that the black box stance, according to Weed, is a "*methodologic approach* that ignores biology and thus treats all levels of the structure

[85] For example, there can be two approaches to predicting whether the turning of the key in a car of the brand *B* starts the engine. We can observe the correlation between the turnings of the key and the engine startings in different cars of brand *A* and then extrapolate the results to predict the engine starting by the turning of the key in a car of brand *B*. On the other hand, we can "lift the hood" and try to understand what makes the type of an engine that these brands share start by the key being turned.

below that of the individual as one large opaque box not to be opened" (Weed 1998: 13, emphasis added). We have seen in Chapter I how epidemiological research on disease causation, because of its statistical approach, takes risk factors rather than causes as the category associated with disease etiology. Similarly, clinical trials and observational studies take anything that happens, for example, between the administration of a treatment (whether it be surgery, a drug, or some physical exercise) to the occurrence of measurable outcomes as the black box. The details within the black box can sometimes be distracting. What matters are the numbers. Epidemiologists are sometimes even explicit about the ignorance and speak in its favor when approaching the study of disease etiology. Again, clinical trials and observational studies are not designed to reveal what happens inside the black box. The designers of these studies and the investigators performing them are usually not concerned with what happens in the black box. They are concerned with how good a correlation between the input or exposure and the outcome is.

The difference between the two approaches as I showed here, and as Howick also points out, is that mechanistic reasoning "involves looking 'inside the black box' at what happens to the relevant mechanisms affected by an intervention", and then reasons about counterfactual scenarios involving the outcomes of a mechanism as a whole or of some of its productive stages (Howick 2011a: 125). Only if understood in this way, can we say that mechanistic reasoning is something different from or opposed to usual EBM's predictive reasoning.

If the black box approach to clinical studies stops at the level of the individual, then the mechanistic approach, on the other hand, goes into the individual. As implied by many authors, this requires different types of studies, as well as different types of evidence. The studies or evidence-gathering methods of the mechanistic approach are not large scale experimental or observational trials. Rather, they include in vitro, in vivo, and often times in silico experiments on biological mechanisms. Oftentimes they include "experimental systems" like resus macaques (Weber 2004, Aronson et al. 2021). The type of evidence required for a prediction claim that *A* will cause *B*, then, is usually considerably different on those two approaches and so are their evidence-gathering methods. Both approaches assume that the black box stands for all known and/or unknown mechanisms. Mechanisms within the black box are not necessarily biological, chemical, or physical mechanisms. When considering public health these mechanisms can also include social and psychological mechanisms. Nevertheless, the discussion in philosophy of medicine is usually concerned with biological, chemical, and

physical mechanisms. Concerning the administration of a certain drug to a patient, Howick refers to the mechanisms within the black box with an abbreviation ADME, common in pharmacokinetics and pharmacology. ADME stands for "absorption, distribution, metabolism and excretion". Therefore, to arrive at the predictive claim via mechanistic reasoning about the outcomes of administering a particular drug to a patient will require taking into consideration various mechanisms and complex interactions occurring in those four stages. The empiricist strategy simply avoids looking into these mechanisms and proceeds with the comparison of outcomes between different groups and then extrapolates the results from the study group to the target group.

To repeat, the claim I defend here is that there is no distinct mechanistic reasoning in addition to well-known types of inferences and reasonings already acknowledged and discussed in philosophy of science or epistemology of science. Mechanistic reasoning is not something that philosophers of science and epistemology, logicians, cognitive scientists, and psychologists have simply overlooked. It is distinguished by what its name implies – the inferences that are used are concerned with the workings and the outcomes of biological and chemical mechanisms underlying the associational or correlational relationships of randomized controlled trials and observational studies.

I will present a more detailed discussion on the structure of mechanistic reasoning in the following section, and especially in section 3.7.1. For now, it is sufficient to understand that mechanistic reasoning just is the inference about or through the productive stages of a mechanism, no matter the particular inferential strategy used to reason about these stages. Mechanisms are, however, always represented by a particular model of mechanism. Therefore, the type of reasoning used when reasoning through a model of mechanism is conditioned by the type of representation or a model of mechanism. If the model is a diagrammatic representation or picture of mechanism that uses boxes, arrows, and spatial arrangement to represent the mechanism, then reasoning through the mechanism will probably involve mental simulation. If the model is represented as a set of equations, reasoning about productive stages will include mathematical modelling, and so on and so forth.

Whether it is concerned with extrapolating results by simple induction or by interpretation of results of clinical and observational trials, EBM uses statistical and probabilistic reasoning found in all sciences that use a statistical approach to understand relations between variables and comparisons between populations (for example, in economics).

The EBM movement is famous, however, for its attempt to grade the quality of both the evidence and evidence-gathering methods, experimental or observational. For example, some randomized controlled trials are performed better than other randomized controlled trials. Also, systematic reviews are graded as better evidence than large scale randomized controlled trials. In other words, the EBM movement imposes constraints and conditions that a high-quality study and high-quality evidence (statistical or populational) is supposed to satisfy. The literature on trial design is vast and elaborate, and it has been discussed in philosophy as well. Aside from succumbing to the general constraints of inductive statistical reasoning, trial design models provide the conditions that experimental study ought to satisfy (e.g., from the elimination of potential biases and positing the null hypothesis to interpretation of data). If studies are of high-quality then their evidence will be as well, and that makes simple induction and extrapolation more trustworthy. Similarly, a philosophical account of mechanistic reasoning should impose constraints and conditions where their satisfaction allows grading the quality of its instances and the evidence that they provide. Is there a way to grade mechanistic reasoning as being of higher or lower quality?

Unfortunately, not much has been said about these conditions, or about the criteria for grading the quality of mechanistic reasoning. A notable exception is a handbook by Parkkinen et al. (2018). Concerning mechanistic reasoning, then, the discussion in philosophy of medicine has mostly been concerned with its role and place in medical science and practice. The usual structure of discussion is the following. After presenting case studies or examples that aim to show the unreliability of mechanistic evidence in assessing prediction claims about the outcomes of interventions, authors seek out other potential uses of mechanistic evidence. This is the line of argument taken in, for example, Bluhm (2011), La Caze (2011), and Andersen (2012). These other uses of mechanistic evidence are usually recognized to have crucial roles in drug design (Mavromoustakos et al. 2011, Aronson et al. 2018), in the interpretation of experimental evidence (La Caze 2018), and in extrapolation of evidence from experimental and observational studies (e.g., in Steel (2008)). Similar arguments are usually presented by the members of the so-called EBM+ circle.

Howick (2011), Solomon (2015), and Aronson in his (2020) discuss a rather different understanding of mechanistic reasoning than the one asserted by the authors from the previous paragraph. They argue that we should differentiate between two aspects of evidence concerning mechanisms. This differentiation was already introduced in the introduction of this dissertation:

evidence of mechanisms and *mechanistic evidence* or *evidence from mechanisms* (that is, mechanistic reasoning). Descriptions of mechanisms are not, by themselves, mechanistic evidence. They become evidence if they allow inferences about some specific set of unobserved causal relations that are relevant for medical treatments. A description of a mechanism becomes evidence when we can infer causal relations involving or going through that mechanism and when we can derive positive patient-relevant outcomes from it.

The knowledge of mechanisms gathered by medicine's basic sciences reveals the puzzling, complex, and interrelated workings of different parts of the human body. Knowing how these mechanisms and processes fail gives us an understanding of the processes that lead to diseases. In this sense, evidence of mechanisms is certainly informative and explanatory for patient-relevant outcomes. It is hard to imagine how we could ever start explaining diseases and consider treatments without any knowledge of the biological causes of diseases and the 'normal' functioning of the human body (although there are episodes in the history of medicine where such interventions were brought about by sheer accident).

But this is not the same as mechanistic evidence. Solomon concludes her discussion by placing mechanistic reasoning in the context of discovery as opposed to the context of justification. The context of justification in medicine is taken by Solomon to be EBM's epistemological framework, particularly the experimental and observational studies at the top of the hierarchies. Evidence of mechanisms is important because, as she claims, "[one] of the tools of discovery is thinking about mechanisms" but rarely do we gather mechanistic evidence from evidence of mechanisms (Solomon 2015: 125). The knowledge of mechanisms (evidence of mechanisms) provides us with grounds for thinking about how we can exploit them for medical purposes (mechanistic predictions). But predictions about medical treatments that are inferred from the knowledge of biological or chemical mechanisms (for example, pharmacokinetic and pharmacodynamic properties of a drug and its measurable effects on humans) should be tested in clinical epidemiological studies, presented in sections 1.5.2. and 1.6.

Certainly, there is a difference between having evidence that a particular mechanism exists or that it is responsible for some phenomenon and being able to successfully predict *from* that evidence *to* the claim that *that* mechanism will give such and such output in yet unobserved conditions. I have stated numerous times that Aronson's distinction between evidence *for* a mechanism and evidence *from* a mechanism is very helpful. These straightforward terms

already allude to an end point that a particular evidence is used for. It seems that a lot of work done in philosophy of medicine tends to confuse those two notions of having evidence about mechanisms. That said, Howick, Solomon, and Aronson are right: the philosophical discussion on mechanisms in medicine should be clear about this distinction. Mechanistic evidence or mechanistic reasoning should be a term reserved for prediction activities based on evidence and models of mechanisms. The interpretation of results of clinical trials and the design of a study are not concerned with prediction claims.

Some medical claims will only require evidence of mechanisms. For example, we know that the infection with SARS-Cov-2 can lead to ARDS. Explaining how this comes to be requires evidence of mechanism(s). In drug research, evidence of mechanisms and mechanistic reasoning (or mechanistic evidence) usually come together and they both figure in the process that resembles developing and testing mechanistic hypotheses (as this has been described in "The New Mechanistic Philosophy" literature). A successful mechanistic hypothesis offers evidence that a model or at least some part of a model of mechanism represents the real thing in a satisfactory manner. Consider finding the appropriate minimum and maximum dosage of a drug for different populations. Thinking about mechanisms involved in absorption and distribution and considering chemical properties of drug compounds provides grounds for inferring hypotheses about drug dosage regimes – prediction claims – which are then tested in clinical trials (since mechanistic evidence can usually only provide qualitative claims). Quantitative prediction claims are rarely if ever available via mechanistic evidence.

Why does this distinction matter? Evidence of mechanisms and mechanistic prediction are two different inferential practices, and it is reasonable to assume (even if we consider prediction as a kind of explanation) that they will have different conditions of implementation and different criteria for rating and assessing their quality. Nevertheless, even if we granted that mechanistic reasoning should only be applied to prediction activities and not to a wider understanding of the use of knowledge of mechanisms in medicine, the characterization that Howick and Solomon discuss still offers a rather narrow view of mechanistic reasoning and its role in contemporary medicine.

There are two reasons why I take this to view of mechanistic reasoning to be narrow, and therefore, potentially misleading for the discussion. The first reason relates to what counts as mechanistic evidence on Howick's definition of mechanistic reasoning, which Solomon, although critical of it, implicitly accepts. In Howick's view, representations or models of

mechanisms become evidence if they allow for making inferences that some treatment will have a specific positive outcome. That is, a model of mechanism only becomes evidence when we can infer causal relations involving or going through the mechanism, which will have patient-relevant effects (where this is understood as claims about medical treatments). The second reason for considering it a narrow interpretation is that this characterization *always* requires interventions, that is, predictions of the outcomes of *interventions into mechanisms*. Hence, the intervention into a mechanism must cause, produce, or bring about the output. The patient-relevant outcome is the intended outcome of the intervention mechanism: it is supposed to cure disease, eliminate symptoms, or prevent the onset of the disease.

Thinking about mechanistic reasoning in terms of medical interventions into a mechanism is common in the discussion. Fuller distinguishes between two models of mechanistic reasoning used for medical treatments, both of which require intervention (for example, new treatment procedures such as a particular kind of surgery or a new drug): *intervention mechanisms* and *interventions into a mechanism* (2016: 103). The first involves a mechanism of intervention that directly produces some patient relevant outcome. The second is concerned with intervening into one of physiological or pathogenic mechanism's component parts (which usually lies at the so-called *bottleneck* where all the pathways in the mechanism converge and lead to the output) in order to obstruct the mechanism from producing the output. Thagard in his (2003) paper distinguishes between medical treatments that either induce or inhibit a certain causal pathway involved in the pathophysiology of a disease (corresponding to Fuller's interventions into a mechanism). Thagard also assumes that some medical interventions "help" a mechanism in producing the output rather than preventing it from occurring.

As I said, I take this to be a rather narrow understanding of what mechanistic reasoning amounts to in medicine. If it is understood as an inference about prediction claims based on models of mechanisms, mechanistic reasoning then refers to any inference about productive stages in a mechanism, in which the end product is a prediction claim about the output of a mechanism as a whole or the outputs of its particular stages. There is no reason to assume that mechanistic reasoning necessarily involves interventions into a mechanism, or that it is strictly about patient-relevant outcomes. Mechanistic reasoning through a certain model of mechanism by mental simulation does not necessarily imply interventions, regardless of whether interventions are understood in the Woodwardian sense or some other sense. Mechanistic

reasoning in medical science and practice, then, is only a particular kind of a more general approach to prediction. In that regard, Howick's and Solomon's views on mechanistic reasoning exclude from the discussion other potential domains of use of mechanistic reasoning. The scope and utility of mechanistic reasoning stretches beyond the prediction of positive patient-relevant outcomes of interventions. Thinking about interventions that cause patient-relevant outcomes does not necessarily seem to be applicable, nor does it necessarily reflect mechanistic reasoning in predictive claims about prognosis or diagnosis. A prognostic claim about the outcome of untreated disease *X*, based on mechanistic rather than statistical reasoning, does not necessarily refer to an outcome of any intervention. I am tempted to say that in cases of prognoses, where a doctor says to a patient that their case of disease *D*, if untreated, will lead to such and such outcomes, the doctor's prediction is assuming a scenario lacking any intervention into disease *D*.

Rather than proposing an encompassing definition of mechanistic reasoning that would potentially be either trivial or too strict, I will conclude this discussion with the following three points that characterize mechanistic reasoning in medicine, and which will serve as starting points for the discussion to follow:

> **MR1.** Mechanistic reasoning is a prediction activity concerning outputs of mechanisms.

First, mechanistic reasoning is not a specific kind of reasoning that is distinct from, for example, deductive, inductive, or abductive reasoning. It is distinguished by the content of reasoning – namely, the biological, chemical, social, or psychological mechanisms rather than the population-level outcomes and laws, and where "mechanism" is understood in the epistemic sense of developed in the previous chapter (a model of causal structure that is comprised of component parts and component operations organized in some specific way that altogether explains the end-product).

> **MR2.** Mechanistic reasoning operates through models of mechanisms, not through "real" mechanisms.

Second, reasoning through a model of mechanism means that the inference is based on the features of models of mechanisms, not on the features of worldly mechanisms. These models include linguistic representations, diagrammatic representations, causal Bayesian nets, sets of equations etc.

MR3. Reasoning through a model is conditioned upon the type of a model of mechanism.

Third, since mechanistic reasoning in medicine is characterized as an inference about the productive stages of models of mechanisms relevant for medical science and clinical practice, in which the prediction claim is its end-product, the quality of reasoning about the expected outcomes of each of the productive stages in mechanisms is determined by the type of model (or representation) of mechanism. Hence, the quality of reasoning is rated by the standards of a particular method used in asserting every prediction claim throughout the productive stages – e.g., deductive reasoning, different types of inductive reasoning, analogy, Bayesian nets, computational modeling, and so on and so forth. This does not mean that deductive reasoning involving models of mechanisms cannot be used in medical science and practice, but it is probably rarely done in the way that Solomon discusses it.

In the next section, I expand the view of mechanistic reasoning beyond MR1, MR2, and MR3. I discuss in more detail what these conditions should amount to in practice by considering that mechanistic reasoning is still a model of prediction based on explanations of the sort discussed by mechanistic philosophers.

3.7. Mechanistic reasoning in interventions and diagnosis

What makes some prediction activity or inference better than the other? An obvious and certainly the most important criterion is that it provides true prediction claims most of the time. If some prediction model regularly produces false prediction claims, then, simply that model performs poorly. The proponents of the EBM movement argue that mechanistic reasoning regularly fails to provide true predictions. Such a claim is acknowledged by the majority if not all philosophers of medicine and science. But some predictions in medicine based on the evidence of mechanisms have turned out to be true. So, when does mechanistic reasoning succeed and why and when does it fail? What conditions should mechanistic reasoning satisfy to be qualified as an instance of high-quality prediction inference? And finally, what does the structure of mechanistic reasoning look like?

I narrow my discussion to two instances of mechanistic reasoning in medicine. First, I discuss prediction of outcomes of medical treatments or interventions into mechanisms. As I

discussed in the previous section, this has been the main point of discussion on mechanistic reasoning in the literature, and it is certainly something that should be of special interest from the medical point of view as well. I approach this issue by discussing first how we should think about or categorize medical interventions and how mechanistic reasoning fails. The structure of mechanistic reasoning and the conditions and/or constraints of a normative account of mechanistic reasoning should come as consequences of locating and explaining the failures of mechanistic reasoning. Second, I have claimed that mechanistic prediction does not only include prediction of interventions of medical treatments but provides a wider scope of claims. This is a consequence of my argument that prediction claims based on the evidence of mechanisms do not necessarily include wiggling with parts of the mechanism in order to see the results nor are they only concerned with treatment outcomes (predictions of future events). In the last section of the chapter, I claim that this happens to be the case in some examples of diagnosis. I argue that when diagnostic claims are based on models of mechanisms, they should be considered retrodictive claims.

3.7.1. Mechanistic predictions of the outcomes of interventions

There are numerous characterizations of interventions into mechanisms for medical purposes. I have already mentioned that Fuller distinguishes between interventions into a mechanism and intervention mechanisms, whereas Thagard differentiates intervening into causal pathways either to inhibit their normal function or obstruct their pathophysiological functioning. Function and dysfunction are at the center of Moghaddam-Taheri's account from (2011).[86] She discusses medical interventions primarily as a way of restoring the regular functions of dysfunctional mechanisms. We have seen that such a consideration relies heavily on a loaded understanding of what the proper functions of dysfunctional mechanisms are, and, additionally, a measure of good faith that such restorations can always be achieved.

These characterizations usually focus only on a certain type or types of medical interventions.[87] But just a cursory glance on medical interventions reveals a variety of different medical treatments that are all considered interventions, one way or another. Administrations

[86] Similarly, I have presented Garson's implication that functionality provides the rationale for thinking about diseases and their treatment in his discussion on functions and mechanisms in (2013).
[87] Fuller's characterization is perhaps the most encompassing.

of drugs are interventions that aim to restore or disrupt a function or dysfunction of some mechanism. Surgeries are performed to remove cancerous tissue, infected tissue in fistulae, kidney stones, or parts of the thyroid gland, and so on and so forth. Interventions are performed to restore broken bones, joints, or raptured ligaments and tendons. Some interventions will include plates, screws, pins, or stents to interfere with pathological mechanisms. But some interventions, on the other hand, are not performed on patients that necessarily have a disease. For example, cosmetic surgeries, hair transplants, or circumcisions are all medical interventions that often do not aim to cure disease or restore the function of dysfunctional mechanisms.

As this shows, medical interventions are a diverse set of treatments and some of the aforementioned characterizations do not consider or encompass many of them. I propose that medical interventions are perhaps better characterized or categorized by the type of mechanism into which they intervene – the type of mechanism model we have chosen to represent a phenomenon. In this regard, I distinguish between three types of medical interventions – interventions into pathogenic, pathophysiological/pathological, and physiological mechanisms. The kind of intervention we will be discussing, in that case, will depend on the goal of medical treatment that we have set out to achieve: to prevent a disease from obtaining, to restore a dysfunctional mechanism, or simply to intervene into a normally functioning physiological mechanism in order to achieve other health related outcomes.

I take interventions that aim at pathogenic mechanisms as directed towards prevention or interference with disease etiology. They are made to obstruct a pathogenic mechanism in achieving its output. Most chronic diseases have a multifactorial etiology and interventions that aim at one of the causal chains comprising an etiology may not turn out to be effective in disease prevention overall. However, most infections have at least one *necessary* cause. The HIV virus is a necessary cause of the HIV infection, and tuberculosis is a disease that arises from contracting *Mycobacterium tuberculosis*. As we have seen these are not sufficient causes, but they are necessary. Preventing the infection of an organism with *Mycobacterium tuberculosis* will result in the prevention of the pathogenesis of tuberculosis. Therefore, I associate interventions on pathogenic mechanisms primarily, but not necessarily, with preventions or treatments of bacterial and viral infections. Here, the intervention aims at some stage of a disease mechanism before or after a bacterium or a virus has entered the organism and started reproducing and interfering with physiological mechanisms' regular workings. That

is, antiviral or antibiotic therapies will aim at counteracting the reproduction mechanisms of viruses and bacteria. For example, most HIV antiretroviral drugs work as inhibitors of reverse transcriptase and protease, two important enzymes in the mechanism of the HIV-1 replication cycle, and in that way prevent the spreading of infection and the occurrence of pathophysiological processes.

Interventions into pathological mechanisms are probably the most common type of medical interventions. They treat symptoms and cure diseases or other negative health-related conditions. They include surgeries, drug administrations, physical therapies, psychological therapies etc. These interventions are performed to restore functions of failed or dysfunctional physiological mechanisms. For example, Andersen in her paper discusses knee lavage and debridement as treatments (surgeries) for knee osteoarthritis. Surgeries of clavicle fractures aim at restoring the proper alignment of fractured bone pieces. ACL reconstructions replace the torn ligament with the patient's own hamstring or other suitable tissue. Defibrillation of the heart is another straightforward example of restoring the function of a mechanism that is somehow damaged or broken. Interventions are also used to overcome the effects of pathophysiological mechanisms. These are interventions that, for example, enhance a causal pathway in a mechanism that is somehow failing or being obstructed. For example, sildenafil citrate aims to inhibit the dissolution of cGMP by binding to the PDE5 enzyme in the NO-cGMP causal pathway, thereby enhancing the output of the pathway by increasing the concentration of cGMP. It allows smooth muscle cell relaxation and enhances blood flow in the corpus cavernosum. Insulin administration substitutes low levels of insulin while stents are used to unblock pathways of, for example, blood vessels due to blood clots. All these interventions are examples of repairing and restoring broken, dysfunctional physiological mechanisms or simply overcoming the effects of pathological and pathophysiological mechanisms.

Medical interventions are not performed exclusively to treat or cure. Interventions on functional, "normal" physiological mechanisms, which are for some reason characterized as unwelcome or where the interference with the normal workings of the mechanism is indirectly positively related to health, are regularly performed in contemporary medicine. I have already mentioned cosmetic surgeries and hair transplants. However, the most common and widespread interventions on regular and normally functioning physiological mechanisms are performed in the majority of anesthetic interventions. Of course, a particular anesthetic intervention will

depend on the patient's comorbidities and risk factors, but a great deal of anesthetic interventions will be performed on "fully functional" physiological mechanisms.

Why categorize medical interventions in this way? What can such a categorization reveal that we were missing by considering other categorizations? I provide three points.

First, this categorization implies that interventions to treat or cure the same disease can be different depending on the desired goal – prevention, treatment, care, or some other health-related goal. The broken-normal distinction, as Moghaddam-Taheri argues, is a widely used and helpful heuristic, but it is not the only one. Some broken mechanisms cannot be repaired, and some interventions do not aim at broken mechanisms. Hence, this categorization implies that models of mechanisms, whether pathogenic, pathological, or physiological, ground the rationale behind the kind of intervention used, that is, how the intervention is conceived, performed, and, finally, how its success is measured.

Second, as noted above, this categorization emphasizes the central role these models of mechanisms play in mechanistic reasoning, as opposed to real, worldly mechanisms. Of course, interventions affect the real-world causal structures but reasoning is not "performed over" these causal structures. The centrality of models of mechanism rather than real, worldly mechanisms, as I have announced in the previous section, is important when trying to understand both why and how mechanistic reasoning fails, and how good mechanistic reasoning is or should be structured.

Before stating the third point, let me first consider how mechanistic reasoning fails. In his (2016), Fuller presents one of the more comprehensive discussions of failures of mechanistic reasoning. He distinguishes between three problems for mechanistic prediction: "The first problem is framed as an issue with our cognitive activities or reasoning, the second as an issue with our knowledge or mechanistic model, and the third as an issue with the worldly mechanisms" (Fuller 2016: 100). Consider the discussion from Chapter II. These problems are related to all three theses concerning mechanisms as they have been described in the previous chapter: the ontological, the epistemological, and the methodological thesis. Both Howick and Andersen covertly acknowledge that failures of mechanistic predictions are due to any of the three problems mentioned by Fuller (e.g., failures of modularity or the causal faithfulness conditions, incomplete models, and the stochastic nature of some mechanisms). Of course, mechanisms can simply fail to bring about phenomena. But to acknowledge that some

mechanistic prediction failed because the mechanism simply failed to work would imply that we are confident enough in our understanding of that mechanism's functioning that we attribute its failure to a "glitch". Rarely do we understand mechanisms that well in medicine.

All mechanistic interventions are based on models of mechanisms. Models of mechanisms represent the mechanisms that we have singled out in accord with their outputs. In other words, mechanisms perform functions, or they are mechanisms for a behavior. That mechanisms are always mechanisms for some behavior is a claim repeated and upheld by the majority if not all philosophers discussing mechanisms. But then it seems that different, functionally identified mechanisms will often share components. Intervention into a component in one mechanism can and often will trigger or change the value of a component within a different mechanism, which can subsequently lead to some unwelcome or unpredicted outcomes. In order for the intervention to produce the outcome we desire, it must sever all pathways leading to the outcome, or at least we have to be sure that we have individuated the *only* mechanism that leads to the output and that our intervention can be performed on *that* mechanism – let us call this "the only one mechanism" condition.

Andersen thinks that the only one mechanism condition always requires the modularity feature of mechanisms, and this presents, perhaps, an insurmountable obstacle. She expresses this doubt in the following passage: "But here is the kicker for medicine: in many cases, if we were to include the causal variables that become relevant when intervening on a specific system but are not a part of the mechanism in normal functioning, we eventually end up including pretty much everything in the body. The bodily mechanisms that malfunction, and on which we intervene in medicine to restore healthy function, are not modularly independent from other causal structures in the body. If we want to add more variables to achieve modularity, then we end up in a situation where the entire organism is the first plausibly modular unit we encounter" (Andersen 2012: 995-996). It can be argued that modularity is perhaps not what Andersen is talking about here but rather the causal interconnectedness of bodily mechanisms. Whether it is due to failed modularity or not, interventions on bodily mechanisms regularly trigger multiple pathways in the body that lead to unforeseen effects. The "only one mechanism" condition, then, is a condition that is satisfied in mechanistic reasoning relatively rarely.

Most of the time, there are multiple mechanisms and causal pathway leading to an outcome, and they interfere with each other along the way, much to our lack of awareness or knowledge. Several mechanisms can interact in the production of some outcome. An outcome

may come through several pathways within the same mechanism. The same mechanism can produce a completely different output depending on even the slightest changes in the input variable. Consider that many drugs have to be taken with other drugs because they have multiple effects throughout the body. Similarly, many drugs cannot be taken if the patient is already on some other medication because their interactions manifest negative health-related effects. This complexity has a different side too. Since different pathological and physiological mechanisms can share multiple causal pathways, it is always possible that a drug designed to cure one disease fails to deliver the desirable outcome but nonetheless produces a positive outcome in patients with a different disease. In such cases, the drug's purpose can be changed in order to treat other diseases that were not initially targeted (recall the example with sildenafil citrate). Changing the purpose of a drug is known as drug repurposing or drug repositioning, and it is becoming or has already become an important methodology for discovering medical treatments.

All discussions of mechanistic reasoning failures in the literature (e.g., Howick 2011a, Howick 2011b, Andersen 2012, Broadbent 2013, Fuller 2016) point out that the problems of mechanistic reasoning are due to models of mechanisms that such reasoning is based upon. Certainly, there are failures in methodology or the reasoning in constructing mechanistic predictions, and some mechanisms turn out to be stochastic or much more complex than we initially understood. But most if not all failures of mechanistic predictions in medicine can be traced back to inadequate models of pathogenic, pathological, or physiological mechanisms. Fuller discusses three ways in which models of mechanisms fail to generate true predictions: "model incorrectness", "model incompleteness", and "the amount of abstraction in our model" (Fuller 2016: 100). In other words, models of mechanisms can represent things that are not there, or they can miss some part, causal relation, or organizational feature that turns out to be important for the mechanism's overall functioning. Even when correct, a model of mechanism is usually abstracted and idealized (either in terms of its parts or their causal relations) to such a degree that predictions about the model's behavior regularly fail.

Finally, let me present the third reason why we should prefer categorizing medical interventions by whether they intervene into pathogenic, pathophysiological, or physiological mechanisms. Such a view avoids relating medical interventions to thinking about the functionality and dysfunctionality of real mechanisms. Whether it is an intervention into a pathogenic, pathophysiological, or physiological mechanism, a medical intervention is then

understood as a causal process or route designed to change the value of a certain output. By medical intervention, we aim to either inhibit or induce a certain *Y* in the causal chain/mechanism/pathway linking *X* and *Y*, no matter how that causal chain, mechanism, or pathway is understood. Whether an output is understood as a phenomenon of a pathological or a broken physiological mechanism becomes irrelevant. Such metaphysical concerns do not seem to change the view that medical intervention, in the end, is only concerned with whether the occurrence or inhibition of a designated outcome is desirable or not for the patient's overall health.

Let us consider this idea along the lines of Woodward's interventionism first. If the mechanism in question is a replication mechanism of a virus or bacterium, the intervention into the value of a variable (e.g., reverse transcriptase, *X* in Figure 11) changes the value of that variable (from 1 to 0, when the inhibitor is attached to it), which leads to the changing of the value of the output variable (*Z* – from 1 to 0, no replication). In Figure 11, causal relations from *W* to *X*, from *X* to *Y*, and from *Y* to *Z* are endogenous. Intervention *I*, then, is an exogenous causal process that severs the causal relationships between *X* to *Z* (if, of course, other conditions of Woodward's interventionism are satisfied).

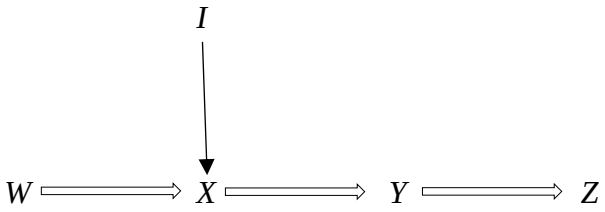


Figure 11. An example of an intervention into a model of mechanism

In Figure 11, a mechanism is represented as a simple directed graph with variables *W*, *X*, *Y*, and *Z*. Models of mechanisms relevant in the biomedical sciences are usually much more complex. However, no matter the kind, models of mechanisms play the main role in mechanistic reasoning about medical interventions. In addition, in most cases the failure of mechanistic reasoning can be ascribed to inadequate models of mechanisms.

But what models exactly? So far, I have assumed and asserted that models of mechanisms, through which mechanistic prediction activity reasons, are models of either

pathogenic, pathological, or physiological mechanisms with interventions designed as exogenous causal processes. Medical treatments are then conceived as interventions that change the value of a given variable within those mechanisms. However, I propose a different view of medical interventions based on mechanistic models.

> **MR4.** Mechanistic reasoning concerning medical interventions amounts to constructing a model of *mechanism of intervention*.

In light of the discussion from Chapter II, I propose to reconsider medical interventions as models of mechanisms themselves. In this regard, mechanistic reasoning about medical interventions is indeed an inference through a model, but through a model of mechanism of intervention (in part based on the knowledge of pathogenic, pathological, or physiological models). Importantly, I am not using the term intervention in a Woodwardian sense here, that is, as the surgical change of a value of some variable within a mechanism. Rather, intervention is here understood as *a mechanism itself*, designed to produce a certain outcome. There is a model of mechanism of blood pumping throughout the body but so I conceive a model of mechanism of regulating blood sugar levels by insulin pumps. What was thought of as an intervention in the interventionist framework, here is just one of the endogenous variables in the causal structure. The example of an insulin pump is particularly vivid, and nicely exemplifies what I have in mind since it includes a machine – the pump – composed of hardware parts and a software that calculates and regulates blood sugar levels. Glucose levels measured by the pump, cause the pump to calculate the required dosage of insulin. There is no reason to doubt that, considering the discussion from the previous chapter, the phenomenon of regulating blood sugar levels by the pump cannot be explained or represented via a mechanistic framework. Here, the mechanism underlies regular behavior since the levels of blood sugar are continuously monitored and adjusted. On the other hand, interventions can be thought of as one-off mechanisms or Glennan's ephemeral mechanisms. For example, the open reduction internal fixation (ORIF) is a type of surgery performed in cases where fractured bone pieces are displaced in such a way that the bone cannot heal properly on its own. The intervention uses screws, pins, plates, or rods to align the broken pieces in order for them to heal. This would amount to Fuller's intervention mechanism. But notice that on this view all interventions in the form of a medical treatment are intervention mechanisms.

For the most part, the discussion on mechanistic prediction in medicine focused on reasoning through mechanisms (from inputs to outputs) in order to infer possible outcomes. I

am not saying this is incorrect. However, the normative account of mechanistic reasoning that I am proposing here turns the picture around. The criterion for good mechanistic prediction of medical interventions ought to be the following: the prediction of intervention mechanisms is structured as a kind of causal explanation of a mechanism that is able to produce a desirable outcome *that has already been determined*: to treat a particular symptom, to cure a particular disease or to achieve any other health related goal. Let me then stipulate a fifth claim about mechanistic reasoning, one concerning predictions of the outcomes of interventions.

MR5. The outcome of a mechanism of intervention is set in advance.

Considering predictions of the outcomes of medical interventions, then, the process of mechanistic reasoning (reasoning through a mechanistic model) should have a similar structure to that of constructing a mechanistic explanation. That is, we are not trying to figure out what the value of variable *Z* will be but rather what kind of causal structure will produce the value *z* of variable *Z*. Hence, mechanistic reasoning is a process of constructing a model of intervention mechanism *for that* outcome. It amounts to reasoning through the productive stages of *that* model and infers not what outcome will obtain but how an outcome that has already been set in advance can be achieved. A mechanistic prediction claim is the claim that *that* mechanism can produce *that* outcome. As I noted above, I am not implying that mechanistic predictions of medical interventions do not necessarily have a determined outcome by the investigators, but I claim that the good ones will have to.

Let me now consider three important characteristics of mechanisms of interventions: their outcomes, models, and methods of reasoning.

THE OUTCOMES OF MECHANISMS OF INTERVENTION

Mechanisms are singled out or identified by their outputs, the regular functions they perform, or simply by phenomena that they account for. Mechanisms, as mechanistic philosophers argue, causally sustain, or maintain regularities. Most mechanistic accounts of explanation in the literature have centered around repeatable or ongoing mechanistic outputs (blood circulation), or mechanisms that sustain regular functions (protein synthesis). However, recall the example of a prototype mechanism from section 2.3.3. - the bomb. If it works properly, it will only work once. The regular function of that *type* of mechanism, I will assume,

is the repeatability of such an output in all of its tokens. In terms of medical treatments, an output, function, or phenomenon of a mechanism is an outcome, function, or phenomenon of medical intervention. In cases where medical treatment does not include a sustained causal operation of its component parts (as in the case of the insulin pump mechanism or bypass), a regular function of a mechanism of intervention, I will assume, is the repeatability of the outcome of intervention in different cases – patients with similar pathogenic, pathological, or physiological mechanisms.

Recall Fuller's claim that inference in mechanistic reasoning works in a similar way as in mechanistic explanation but, "what differs are the components that are taken for granted" (Fuller 2016: 103). In his view, mechanistic prediction takes a certain model of mechanism for granted and reasons through that model to come up with its output. Mechanistic explanation, he claims, works the other way around. In case of constructing a mechanistic explanation, the outcome or phenomenon is defined, distinguished, or clearly characterized. I argue, however, that a good mechanistic reasoning concerning medical interventions has to work the same way as mechanistic explanation – once a certain outcome or value of an outcome variable is set, a model of mechanism that can produce it is developed. I take it that just a cursory glance at medical practice supports this view. We want to raise blood sugar level to a certain point. We want to put the pieces of a broken bone in some alignment so it can heal properly. Often, but not always, instead of questioning what the outcome of a certain mechanism (intervention) will be, medical scientists want to know whether a *certain patient-positive outcome can be achieved*. In the case of sildenafil citrate, a clearly defined outcome – penile erection in conditions of sexual arousal or activity – led scientists to construct a model of mechanism that can produce this effect. This proper characterization of the outcome or effect also led to the characterization of clearly identifiable measures of success of achieving the outcome – measurements of magnitude, timing, and duration of erectile response in particular conditions. In the case of antiarrhythmic drugs developed in the 1970s, the outcome was also clearly defined – reduction in frequency of ventricular extra beats (VEBs) – since it was believed that this will reduce mortality. Andersen's example of knee lavage and debridement also fits within this picture. It is a model of intervention mechanism developed to achieve a certain outcome – restoring the function of a damaged knee mechanism by removing damaged cartilage, bone, and the debris from the joint. The same applies to all previously mentioned cases of mechanistic reasoning.

Therefore, mechanisms in explanations of phenomena and mechanisms in my account of mechanistic predictions of the outcomes of medical treatments are functionally identified. When scientists changed their point of interest from angina pectoris and hypertension to erectile dysfunction, the outcome of intervention has already been clearly defined – the erectogenic effect. To reiterate, the question of good mechanistic prediction, I claim, is not what the outcome of administration of sildenafil in certain conditions in middle aged men will be but whether sildenafil can, and to what extent, produce a certain outcome in certain conditions in middle aged men with psychogenic or organic ED. Therefore, the first criterion for a good mechanistic prediction of medical intervention I propose here is a *clearly defined outcome of intervention* – an effect we wish to achieve with our intervention. It is also the first step in developing a model of intervention mechanism. Of course, I recognize that mechanistic prediction claims about the outcomes of interventions are usually qualitative and not quantitively precise. A mechanistic prediction claim about the erectogenic effect of sildenafil probably cannot specify timing, duration, and magnitude of the erectile response in particular patients. However, the question is how accurate any prediction can really be in this respect? Empiricist predictions are applicable to a specific patient to the extent that patient is similar to the population average or mean. A mechanistic prediction claim, then, should at least be able to give the upper and lower values of an outcome variable.

There are further constraints that mechanistic prediction needs to satisfy. For example, antiarrhythmic drugs reduced the frequency of VEBs, but studies have shown that they, in fact, increased mortality. This case suggests, as Howick has discussed, that there is an important difference in defining and measuring surrogate outcomes and patient-relevant outcomes. An additional criterion of quality of mechanistic prediction and a constraint on choosing the outcomes should therefore be the characterization of (if possible) the undesired side effects of the intervention or the future effects of the outcome itself. We want antiarrhythmic drugs to reduce the frequency of VEBs but not at the expense of increased mortality by further weakening a heart compromised by heart attack.

MODELS OF MECHANISMS OF INTERVENTION

Recall the discussion on the relation between explanation and prediction in the mechanistic literature. It has been claimed that predictions of interventions grounded in the

evidence of mechanisms often failed because models of mechanisms were incorrect or incomplete. But considering the many cases where mechanistic reasoning failed, it can also be argued that it failed because of the inadequate models of intervention, not because of incomplete or incorrect models of pathophysiological or physiological mechanisms. In other words, those models of intervention mechanisms could not produce the desired outcomes, or they produced the desired outcomes, but have also caused some undesirable counterindications (or, as in the case of antiarrhythmic drugs, even increased mortality). A failed prediction in this sense indicates a failed, incorrect, or incomplete model of intervention. In this regard, take notice that in some cases of failed mechanistic prediction, our understanding of pathological and pathophysiological mechanisms (for example, a damaged knee due to arthritis) has not changed (at least not substantially). The model of a disease mechanism and its normally functioning counterpart remains the same. Even though surgical intervention into a damaged knee is perhaps no longer considered to be a more successful medical treatment than physical therapy, our understanding of an arthritis induced damaged knee is still preserved. As mechanistic philosophers claim failed predictions indicate failed models. This is indeed true, but notice that according to this account, the failure is of a model of medical intervention, not of a pathological or physiological mechanism. Therefore, our thinking about those models of mechanisms of interventions has changed – something about these models was incomplete or incorrect. These models of mechanisms did not produce phenomena we thought they would. Hence, the second step in developing a model of intervention mechanism, is the construction of a model of mechanism that can produce the designated therapeutic outcome but which does not lead to undesirable consequences.

How should a model of mechanism of intervention look like? Of course, it should satisfy all the usual conditions and constraints of mechanistic explanation and share the same features with the usual models or representations of mechanisms. First, the three necessary features of a model of mechanism have to be identified: component parts, component activities or interactions between component parts, and its specific organizational features. Second, the construction of a model presumes the identification of productive stages in the mechanism of intervention, from the input to the output. A series of productive stages and their relations should be understood to the extent that their operations are explainable and predictable with as few black boxes as possible. In the case of drug administration, developing a model of intervention mechanism will include all the stages involved in ADME - absorption,

distribution, metabolism, and excretion. These stages should be mechanistically explainable and tested in vitro and later on in vivo.

Some mechanisms work only in suitable environments or when some conditions are satisfied. Stipulating the conditions under which the interventions work should be one part of mechanistic reasoning in these cases. It was suspected that sildenafil could work only in conditions of sexual arousal and sexual activity since in different circumstances there would not be a level of concentration of NO high enough to produce the amount of cGMP that would allow sildenafil to exert its effect. Once the outcome was defined, scientists proceeded to construct a model of intervention mechanism – a model of mechanism able to produce the effect in these circumstances. These were specified after the model had been put to test and they were later implemented in the trials.

Although I am arguing that mechanistic explanation and prediction have the same structure, they are, in fact, different inferential activities. Mechanistic prediction of medical interventions will require additional conditions, ones that mechanistic explanation usually need not to consider. Here I present one such condition.

First, remember "the only one mechanism" condition discussed earlier. It states that an intervention must either sever all pathways leading to an outcome, or we have to identify the *only* mechanism leading to the output. There are two possible ways of understanding the only mechanism leading to the output. The first is that our model of intervention constitutes the only causal pathway by which the outcome can be achieved. The second is that all pathways connected to the outcome, which the intervention can, or does trigger are identified. Whatever the understanding of the only mechanism is chosen, it is meant to ensure that our intervention, to the best of our knowledge, does not lead to other outcomes, especially those that have negative health-related effects. Finally, all these pathways leading to the outcome that are directly causally influenced or indirectly triggered by an intervention have to be included within the boundaries of a model of mechanism of intervention.

How much detail do we need to include in a model of mechanism of intervention to be able to predict the outcome? Explaining how the mechanism works and predicting its outcomes, I presume, will not always require the same amount or the same type of detail. We have seen in the previous chapter that some authors argue that the level of detail required in mechanistic explanation is conditional upon the phenomenon itself and the purpose of its

explanation. Emphasis on detail can sometimes blur the understanding of a mechanism's overall operation. On the other hand, others have claimed that the more details our model includes the better off we are in terms of understanding how the mechanism brings about its output. Remember from the general discussion on prediction at the beginning of the chapter that some philosophers have also claimed that an emphasis on details in explanations can impair our predictive capabilities (for example, in Carrier 2014). So how much abstraction and idealization of a mechanism's parts, activities, interactions, and organizational features can a model of intervention mechanism afford?

In constructing a mechanistic explanation, it is not always crucial or necessary to understand a component's causal interactions beyond its role within a mechanism of interest. Recall that this rationale constitutes the core idea behind perspectivalism discussed in the previous chapter. For example, NO has functions in mechanisms in both the central nervous system and the peripheral nervous system. In explaining vasodilation, the model of mechanism underlying smooth muscle cell relaxation does not include the role NO has in, for example, the maintenance of synaptic plasticity. However, it seems that in developing a model of intervention mechanism, an emphasis on such details can and often will matter. Every productive stage or interaction between two molecules can trigger some unforeseen molecular pathway that can disrupt the achievement of a desirable outcome or lead to some other unwelcome consequences. It is crucial for a model of mechanism of intervention that *the only one mechanism* leading to the outcome with all of its branching pathways has been identified. If molecular pathways diverge somewhere after the input stage and can have negative health-related effects, then we have not identified *the* mechanism of intervention. So, by definition, a successful intervention mechanism necessarily lead only to positive patient-related outcomes.

Although my argument states that medical interventions are to be considered models of mechanisms themselves and that the structure of mechanistic prediction is similar to mechanistic explanation this does not deemphasize the central role of pathogenic, pathological, and physiological mechanisms. In the end, interventions mechanisms are designed to interfere with and alter *these* mechanisms or to restore *their proper function*. In developing models of intervention, then, scientists regularly use both evidence of mechanisms and evidence from mechanisms. Previous studies suspected and later in vitro experiments confirmed that PDE5 can be found in smooth muscle cells in corpus cavernosum. Later research found that it is abundant in platelets and some neuronal cells. Other research focused on revealing properties

of sildenafil and other PDE5 inhibitors as well as properties of PDE5 that contribute to its high affinity for interaction with those inhibitors – evidence of a mechanism. These have inspired new ideas about other possible mechanisms of intervention where sildenafil could be used – evidence from a mechanism. Hence, mechanisms of interventions are always understood with pathogenic, pathological, and physiological mechanisms in the background. To paraphrase Craver and Darden, the nature of the outcome shapes the process of modelling a mechanism of intervention. This means that the type of mechanisms considered will depend on the nature of the disease or the type of the outcome one wants to achieve. The admission of sildenafil in cases of organic ED treats the failing NO-cGMP causal pathway. Surgeries, such as the aforementioned ORIF, aim at restoring pieces of a broken bone to their proper alignment. Antiviral drugs aim at pathogenic mechanisms connected to viral replication. Models of intervention will, then, differ depending on the type of mechanism they intervene into in several respects. But, most importantly, the difference will be reflected in the details of the model, that is, the level of abstraction that allows for assessing the true predictions of their outcomes. A model of arthroscopic intervention will consider far fewer details about component parts, activities, interactions, and organization than it would be needed for models of mechanism of action of antiviral drugs. So, why was the prediction claim about the efficacy of antiarrhythmic drugs bad or false, whereas the claim about the defibrillation of a stopped heart is of better quality? Simply, the defibrillation claim requires fewer details about parts, interactions and organization and therefore allows for a higher degree of abstraction and idealization in order to produce true predictions.

This example leads us to another distinction between models of intervention mechanisms and models of mechanisms in mechanistic explanations, and the last one I will mention here. While mechanistic explanation stops at the outcome or phenomenon, a good mechanistic prediction does not. As said, all productive stages in mechanisms of interventions should be identified and understood. The pathways that diverge somewhere along the path from the input to the output of the intervention should be identified. However, a good mechanistic prediction will have to ensure that the outcome itself does not lead to other negative health-related outcomes. Removing the thyroid gland altogether in order to achieve positive health-related outcome, can, in fact, bring about different negative health-related effects, such as weight gain or weight loss. Of course, whether the outcomes themselves present such risk factors are, and definitely should be, studied in randomized trials or observational studies.

Nonetheless, good mechanistic prediction, if we want to have one, has to consider all the known or possible negative effects of the outcomes themselves.

The regularity of being able to establish models with such an understanding or the guarantee that they do not require clinical trials is, of course, disputable. But if clinical trials are unavailable for different reasons (for example, ethical, as in the case of defibrillation), this should be the standard of a good mechanistic prediction.

REASONING THROUGH MODELS OF MECHANISMS OF INTERVENTION

Finally, what does reasoning through a model of mechanism amount to? I have claimed throughout this chapter that it is not an idiosyncratic type of reasoning. Rather, it is determined by the type of model of mechanism. I will now discuss this in more detail.

Most models discussed in the mechanistic literature represent mechanisms linguistically, in pictures, or as diagrams. Linguistic representations can specify a lot of details in narrative form, and they offer sequential reasoning through mechanisms or the sequential understanding of a mechanism's production. That is, linguistic representations tell a linear story of the productive stages of a mechanism. On the other hand, diagrammatic representation is more common in both biological and medical textbooks. Diagrams and pictures offer different and often more useful approaches to thinking about the mechanisms' parts, relations, organization, and mechanism's overall workings. One can focus on different regions and parts of a mechanism at a time, or reason in a reverse temporal direction. Thinking through diagrams "requires that the scientist engage in mental activities (especially mental simulation) that are rather different from formal deductive inference" (Bechtel 2011: 538). But reasoning about and through mechanisms does not involve only mental simulations. For example, Sheredos et al. discuss these different types of reasoning through models of mechanisms in their (2018): "The graph takes advantage of spatial cognition, whereas the logarithmic equation makes explicit a very precise claim that can and has been challenged (e.g., by those who argue for a power function). Scientists move deftly between linguistic descriptions, diagrams, and equations when all are available, using each to its best advantage" (Sheredos, Burnston, Abrahamsen and Becthel 2018: 933, 934). Mental simulation is a powerful tool but probably not used so much in high quality medical research. In fact, it is quite possible that in basic medical sciences

nowadays, mental simulation of going through the mechanism is rarely used, whereas maybe, it is perhaps more common in everyday clinical practice.

So, what are the criteria of high-quality reasoning through a model of mechanism? Models of mechanisms are diverse. We can represent mechanisms in numerous ways, not just linguistically or diagrammatically. Whatever scientists find convenient or helpful to represent mechanisms is used to represent mechanisms: real 3D models, pictures, sets of equations, causal Bayesian nets, videos, etc. Furthermore, a single mechanism can be explained by means of different representation. Craver and Kaplan, as we have seen in section 2.6.3., argue that a single mechanism is often grasped or understood only if it is represented by several different models, all focusing on one or more, but not all aspects of that mechanism. That is, one model will add something missing from another model of the same mechanism. The more we can model a mechanism, the more we can understand all its characteristics. Thus, it should not be controversial to state that the way in which one stage of a mechanism activates, produces, or interacts with another stage of a mechanism can and will involve different inferential methods. Some of these methods may include traditional types of inductive reasoning, some may include analogical thinking, some may work as inferences to the best explanation, and some may use mathematical modeling. For example, Darden lists some of the strategies used to infer prediction hypotheses in her (2006): "reasoning by analogy", "reasoning by postulating an interfiled connection", "reasoning by postulating a new level of organization", "reasoning by invoking an abstraction", "reasoning by conceptual combination", and "abductive assembly of a new composite hypothesis from simpler hypothesis fragments" (2006: 216). Therefore, I claim that some distinctive type of reasoning through a model of mechanism does not exist. There is nothing special about reasoning through a causal Bayesian net or a diagram, other than that they both represent a mechanism. Hence, the choice of representation or model of mechanism determines how reasoning through that representation or model is manifested. Consequently, the same applies to grading the quality of reasoning. The kind of inferential activity we use to reason through a model is graded by means of its own standards, as long as we think it is convenient or correct to represent causal relations and productive stages within a mechanism in that way.

Evidence-gathering methods such as in vivo and in vitro testing have been characteristic of the rationalistic approach in medicine. However, recent times have seen the enormous advances in various in silico methods (already mentioned in section 2.6.3.). Consider

this passage by Hopkins in a review article on in silico methods in light of Woodward's account of the mechanistic view of causation and the evidence of causation quoted in section 1.5.:

> Various in silico methods for predicting the pharmacological profile of drugs are in development, the most well-known of which is to 'dock' the three-dimensional structure of a compound virtually into the structure of a protein. But among the limitations of docking methods is the need for high-resolution X-ray crystal structures of proteins. These are particularly difficult to obtain for membrane-bound proteins, which account for 60% of drug targets. An alternative approach has therefore been developed that does not require protein structures. This approach works by analysing the chemical structures of ligand molecules that are known to bind to drug targets, to identify the structural motifs responsible for the binding.
>
> Hopkins 2009: 167

The quote from Hopkins expresses word-for-word Woodward's definition of mechanistic causation or, in his terms, geometrico-mechanical causation: mechanistic thinking about causation implies "one can just "read off" which causal relationships are present from geometrical or mechanical properties" (Woodward 2011: 413). Drug repurposing uses the evidence of properties of drug compounds, their targets in biological pathways or mechanisms, their chemical structures etc. In his review article Park (2019) presents some of the strategies used for drug repurposing such as "Knowledge-based repurposing", "Target-based drug-repurposing", "Pathway-based drug-repurposing", "Target mechanism-based drug-repurposing", "Signature-based repurposing", and "Phenotype-based repurposing" (Park 2019: 60, 61). All of these are backed by biomedical data, such as "microarray gene expression signatures, pharmaceutical databases, and online health communities" (Jarada et al. 2020: 2). These could not be of use if modern medicine had not embraced and started using computational methods such as data mining, machine learning, and network analysis. Should we take these methods to be methods of mechanistic reasoning? If mechanistic reasoning is defined along the lines of MR1, MR2 and MR3 then, yes, there is no reason why they should not be considered methods of mechanistic reasoning. Furthermore, if scientists think that such representations are adequate to represent features of parts, causal relations, and the organization of mechanisms of intervention and their target systems, the quality of such reasoning used in those methods is then graded by the standards of the methods themselves.

3.7.2. Mechanistic reasoning in diagnosis

When you pay your doctor a visit, usually, but not necessarily, the first thing your doctor tries to do is to establish a diagnosis. You and your doctor will most likely go through your medical history. The doctor will take notice of the symptoms you have experienced and examine you to reveal other symptoms and signs that usually co-occur with the ones you have already mentioned. Perhaps certain tests will be recommended and performed if found necessary. We have all been through this process at some point. But what exactly is a diagnosis and what is its function?

Philosophers have not paid much attention to diagnosis but sporadic views and discussions do exist (mostly in philosophy of science, but also in epistemology and even logic). When diagnosis and inference involved in it finds itself in the focus of philosophers and the philosophically inclined work of medical scientists and practitioners, claims that diagnosis, in a way, resembles criminal investigation are frequent. It has been asserted that doctors engage in an investigation to find the "perpetrator" by any means available, that is, to determine why and how the patient's symptoms and signs occur. Hence, diagnostic inference, as some of these authors claim, is perhaps best conceived as "a retrospective, narrative investigation that more nearly resembles investigation in history or economics than experiments in microbiology or chemistry" (Montgomery 2005: 57). On the other hand, some have claimed that diagnosis perhaps should be discussed and analyzed within the context of discovery, where, it is implied, there are no strict rules of logic (deductive or otherwise) (e.g., in Whitbeck 1981).

Indeed, the practice of diagnosis includes a versatile set of inferences: from doctors' own clinical experience to using big medical datasets. Stanley and Campos describe the diagnosis in four inferential steps. "Abduction—generating hypotheses to explain observed or experienced events—is the first stage of inquiry. The second stage is deduction: deriving the testable consequences of the hypothesis so that experimental tests can be conducted. The third stage is induction: actually testing the consequences of the hypothesis and using appropriate methods—for example, statistics—to ascertain the weight of the evidence in favor or against a hypothesis. These are stages of a continuous process of inquiry, and we may move in various ways: for example, abduce, deduce, realize that the consequences are untenable, scratch the hypothesis, go back to abduction" (Stanley and Campos 2013: 302, 303). Abduction is assumed as the first step in the process of diagnosis in the EBM textbooks as well. For example, in Guyatt et al. (2015), Richardson and Wilson write: "One can label the best explanation for the

patient's problem as the leading hypothesis or working diagnosis" (2008: 214). However, many authors take that at a general level, diagnostic practice is composed of just two different approaches that combine or encompass different inferential practices (deductive, various inductive, and analogical reasoning) – analytical, probabilistic approach and pattern recognition (for example, in Stanley and Campos 2013, Guyatt et al. 2015, and Reiss and Ankeny 2016).

The probabilistic approach relies on EBM's approach to evidence from clinical research. Guyat et al. describe such an approach as the following. First, it starts with the list of potential diagnoses (hypotheses) based on mechanistic knowledge, evidence of clinical research and/or experience. The second step ascribes the probabilities to each of the hypotheses so that their sum equals 1. The third step consists in performing tests and gathering evidence. New evidence raises or lowers prior probabilities of hypotheses until we find the best answer to the etiological question of symptoms and signs. Patterns can be thought of as characteristic outputs of certain diseases. Pattern recognition, then, is usually linked to analogical thinking and so can be thought of as closer to categorization than explanatory practice. Let us assume that certain pattern exhibited by a particular patient's symptoms and signs is assigned to or recognized as part of a pattern of disease mechanism described in medical textbooks and scientific articles. But in that case, pattern recognition can be a tricky business. Once again, patterns are based on standard textbook models of either pathogenic or pathological mechanisms. These models of mechanisms, however, are highly standardized and idealized representations of physiological, pathogenic, and pathological mechanisms. As Simon argues, chapters in medical textbooks "contain the mixture of abstract model descriptions, theoretical hypotheses, and simple real-world descriptions necessary for doctors to learn to care for their patients" (Simon 2008: 360). Furthermore, symptoms and signs do not always fit the usual patterns of a disease. This makes their classification or identification elusive. But there is more to pattern recognition than merely a process of recognizing or linking the pattern in an actual patient with the paradigmatic case from textbooks and guidelines in everyday clinical practice. Another analogical approach to diagnosis is *Case-based reasoning* (CBR). CBR is a problem-solving approach based on the analogy of solutions taken from previously solved problems. Such an approach is probably discussed more in machine learning and AI, but it has certainly found its place in medical practice too.

With all this in mind, a diagnostic claim is about particulars or token phenomena, and not about types of phenomena or populations. It is always concerned with some specific physiological, pathological, or psychological state of a particular patient. Even when subsuming a particular phenomenon under some type or class of phenomena, such as lung cancer, ARDS, diabetes, or bipolar disorder, diagnostic claim is a claim about a particular patient and their own token lung cancer, ARDS, diabetes, or bipolar disorder, with presumably, their own specific pathological or pathophysiological state. Most if not all diagnostic tests and methods by which we gather evidence have been clinically trialed. Their accuracy, therefore, is based on probabilistic, statistical data of populations (with a frequentist interpretation). But, in the end, an individual patient is supposed to be the target of diagnostic tests in everyday clinical practice.

So, what is the goal that this investigation is trying to achieve? As far as the literature in philosophy and medicine goes, there are multiple views on what is or should be the goal of diagnosis and, consequently, what diagnosis itself is. Medicine, as I noted throughout the thesis, is an applicative science. The ultimate goal of medicine is to treat symptoms and signs, and to cure or prevent diseases. Following this, Whitbeck assumes that prevention or treatment must prevent or treat *causes* of diseases or symptoms and signs. In that regard, she claims that diagnosis is a necessary step in achieving this goal in practice. That is, "[d]iagnosis is the process of inquiry aimed at discovering *the causes* and *mechanisms* of a patient's disease insofar as this information is needed to inform treatment and management decisions to achieve the best medical outcome for the patient, and to prevent the disease in others" (Whitbeck 1981: 324). Similarly, Stanley and Campos claim that whatever diagnosis may be, its role is to arrive at treatment decisions: "We classify disease states as convenient methods to offer therapy" (Stanley and Campos 2013: 301). Therefore, for them, no medical treatment can start without a diagnosis, no matter whether it is true or false, informative, or lacking in detail and certainty. As they explicitly state: "Our logic is: diagnosis first, then treatment" (Stanley and Campos 2013: 300). But, on Stanley and Campos' view, diagnosis is not necessarily a causal investigation. Diagnosis, then, can be viewed as a "bridge" that connects pathological or pathophysiological structures in the patient's body (or just the list of symptoms and signs) to some systematized ways of treatment (guidelines, for example).

Stanley and Campos's view, as well as Whitbeck's, seems to assert diagnosis as a necessary step in arriving at treatment decisions. Nevertheless, this permissive view on what

diagnosis is does not accurately describe all the ways in which diagnosis works in clinical practice. Perhaps this is to be interpreted as a normative view on diagnosis, but surely it is not descriptive. Doctors sometimes prescribe a treatment even if a diagnosis has not been established at all. That is, a reverse story is not at all uncommon in everyday clinical practice: successful treatment itself provides grounds for making a diagnostical claim. For example, the rationale can be the following: since the admission of treatment *T* to patient *P* relieved the patient's symptoms and signs, it must be that the patient's symptoms and signs were caused by disease *D* for which *T* is a symptoms-and-signs relief factor. If diagnosis is to be interpreted as an inferential or epistemological activity by which we are finding out about the causes or mechanisms responsible for the symptoms and signs identified in the case of some individual patient, does it, then, by following arguments from Chapter II of this dissertation, imply that diagnosis is perhaps the activity of providing an explanation of particular symptoms and signs in a particular patient?

At first glance, this seems like a reasonable proposal. Regardless of whether the treatment has been provided with or without diagnosis, diagnosis seems to be a claim about the causes or mechanisms producing a patient's symptoms and signs. Furthermore, if this is true, then, as we have seen throughout the previous chapter, knowing the causes and underlying mechanisms of phenomena either provides grounds for coming up with explanations or just is the explanation itself. Indeed, this view on diagnosis is perhaps the most popular in philosophy of medicine. Schwartz and Elstein explicitly claim this: "The diagnosis is thus an explanation of disordered function, where possible a causal explanation" (Schwartz and Elstein 2008: 224). Thagard thinks the same way about diagnosis and adds a further claim. He states that diagnosis is an explanation of a particular phenomenon or process(es) in a particular patient: "When a patient goes to a physician with a set of complaints and symptoms, the physician's first task is to make a diagnosis of a disease that explains the symptoms" (Thagard 1999: 20). Diagnosis, then, is a causal explanation linking particular symptoms and signs with their putative cause(s). That is, causes of disease or the disease itself explains the occurrence of symptoms and signs, and by that it is a causal explanation of a singular event involving some kind of actual causation.

I argue that this is a false account of diagnosis: a diagnostic claim is not an explanation. I offer two arguments to support this – one is rooted in medical practice and science, and the other is philosophical. I start with the medical one, while the philosophical argument will be addressed further below.

The first argument, it should be noted right away, does not claim that diagnosis is not an explanation, but rather, that diagnostic claims cannot always be explanations. Quite simply, just a cursory glance at medical practice reveals that diagnostic claims are sometimes assumed and asserted without any identification of causes or mechanisms that explain symptoms and signs. That is, diagnosis can be a matter of simple pattern recognition of symptoms and signs, but which lack any knowable cause or mechanism. Numerous cases confirm that the best treatment for such a pattern of symptoms and signs is treatment *X*, and this probably helps in hypothesizing about the underlying cause, but in the end, the identification of a cause need not or cannot follow in every case. In such cases, then, categorization can be viewed as a simple identification, or better yet, subsumption of symptoms and signs under some disease category. For example, consider a common chronic skin disease – rosacea. Its symptoms are severe redness in the face or facial flushing and erythema. Causes of rosacea and much of its pathology remain unknown. It is suspected that vascular hypersensitivity and excessive endothelial stimulation of cutaneous vasculature are common in people with a rosacea diagnosis but how this is caused and how it causes characteristic symptoms remains unknown. For now, the disease is characterized by its symptoms and not the underlying pathology. Diagnosing rosacea then amounts to a recognition or identification of its common symptoms and signs and their categorization under a certain disease label. But this does not amount in any way to an explanation of those symptoms and signs. There are numerous cases like this in medical practice. In such cases, the diagnostic claim about a certain disease in a particular patient does not necessarily lead to or constitute an explanation for that patient's symptoms and signs. Here, diagnosis is considered as categorization.

However, symptoms and signs can sometimes also evade categorization under some disease. As I have noted earlier, Whitbeck claims that diagnosis does not have a goal separate from the general goals of clinical reasoning. The goals of clinical reasoning, as she understands them, amount to "providing the prevention and treatment for disease that will result in the best outcome for the patient" (Whitbeck 1981: 321). Furthermore, she argues that diagnostic inferential activities do not pick out disease names or identify a certain category that the pathological state within a patient can be related to and identified with. Indeed, diagnosis does not necessarily identify some specified disease. It is not at all uncommon for an individual patient's symptoms and signs to be idiosyncratic. Doctors can presume that the patient may have several diseases, the effects of which are overlapping, but still refrain from asserting categorization. Diagnostic tests can indicate different diseases, and, as noted earlier, successful

treatments can sometimes corroborate one diagnostical hypothesis over another, but it is far from certain that symptoms and signs along with successful treatments will lead to an identification or categorization of a specific disease in all cases. Diagnosis often is but need not necessarily be either categorization or identification.

By acknowledging all that has been said, Maung (2019) presents a more generous view on what diagnosis is and what its claims are. He takes diagnosis to be as an inferential activity with no particular clearly defined goal. It can have numerous roles or functions in medical practice, in addition to classificatory and/or explanatory roles and functions. Notably, diagnosis can facilitate testable causal hypotheses, explain patient data, guide possible interventions in terms of treatment and management, organize different features exhibited by a patient in a unified phenomenon, categorize different disease states etc. No matter the approach taken to infer a diagnosis, claims of diagnosis, in the end, can serve as an inferential ground for all of the functions mentioned by Maung, but as I claim below, they themselves are not any of the things included on Maung's list of functions.

The first argument against diagnosis as explanation is that we cannot always explain symptoms and signs by referring to their causes or mechanisms. However, the argument does not say that diagnosis is not an explanation. So, what is the philosophical backing for this claim that I have announced above? As Maung notices, a diagnostic claim can have multiple consequences or inferences drawn from it, of which categorization or explanation can be just one among many. When a certain disease has been characterized and categorized, it can have numerous social and psychological consequences for the person who have been diagnosed with it. But we should not commit a fallacy and mistake diagnosis for its consequences or a further set of inferences that can be drawn from a particular diagnostic claim. Diagnostic claims can indeed serve as grounds for explanation of a patient's symptoms and signs, but they do not amount to explanations. Why?

Many biological and biomedical phenomena, either as regularities or one-off events, as we have seen in Chapter II, are explained within the mechanistic framework. In that chapter, I have discussed the main metaphysical, epistemological, and methodological theses of mechanistic philosophy. Together, these theses amount to a comprehensive philosophical framework for thinking about phenomena, explaining phenomena, and doing scientific research in a particular manner. Most if not all mechanistic philosophers agree that mechanisms are either causally responsible for a phenomenon, in terms of a series of productive stages

connecting inputs and outputs, or constitutive of the phenomenon (that is, some mechanisms just are the phenomenon). This distinction is both a metaphysical and an epistemological thesis (but with inevitable methodological consequences and problems, such as inter level causation). Pathogenic mechanisms are perhaps best understood as etiological mechanisms. They connect exposures with the occurrence or onset of the disease. On the other hand, pathological and pathophysiological mechanisms seem to be easier to conceive as constitutive of the disease. They *are* the disease. Bacterial growth or reproduction (for example, *Mycobacterium tuberculosis*) in some part of your body seems to be just that – the infection. I have also claimed throughout this dissertation that mechanistic explanation amounts to a representation of the phenomenon of interest as a model of mechanism, in which models can assume different forms. A model of mechanism, in the end, should include all three aspects of a mechanism (component parts, activities/interaction, organization) together with a stipulation of input condition in order to explain the phenomenon or output of a mechanism.

A certain disease will manifest certain symptoms or signs, usually characteristic to it. For example, a persistent cough for more than three weeks, chest pain, fever, fatigue, and weight loss are usually symptoms and signs of active tuberculosis. But these are not the disease itself – the infection. On the other hand, the presence of characteristic nodular granulomatous structures in your lungs called tubercles is constitutive of a certain pathological state – active tuberculosis. Reduced ability of erectile response in appropriate conditions seems to be a symptom and sign of ED, regardless of its etiology. But the disease itself just is the pathological mechanism. Having systolic blood pressure over 140 mmHg and diastolic blood pressure less than 90 mmHg is indicative of having isolated systolic hypertension. It is a sign of a certain pathological mechanism, for example, reduced arterial elasticity. Symptoms and signs, therefore, are indicative of the presence of some pathological mechanism because these just are the mechanism's outputs. These are phenomena of pathological mechanisms, in the same way as blood circulation and thump noises are phenomena caused by the heart mechanism. But outputs, outcomes, or phenomena should not be viewed as constitutive of the disease (the mechanism).

So, what do these two paragraphs assert? Diagnostic claims have a considerably narrow scope and content than explanations, especially when considering that mechanistic explanations are models of mechanisms. Diagnosis *does not propose* a model of any kind. It does not arrange parts, activities and interaction, and organization into a model of mechanism.

In that regard, I argue that diagnostic claims, in cases where they are not pure categorization but rather based on the knowledge of a type (rather than token) of mechanisms, fall short of being mechanistic explanations (or explanations of any sort in general) of symptoms and signs. No matter the approach taken (probabilistic based on clinical studies, or pattern recognition based on clinical experience, CBR, or mechanistic evidence), when diagnosis is a claim about a cause or mechanism responsible for symptoms and signs, rather than categorization, it is a type of a prediction claim. More specifically, I take it as a mechanistic retrodictive claim.

Recall that prediction, just as explanation, can be understood in different ways. Specifically, I have claimed that there are two general views on the relation between prediction and explanation in science in the philosophical literature: prediction is a kind of explanation and prediction is a hypothesis about future or past event inferred from a particular explanation. I have also claimed that, depending on the final goal of our inquiry or interest, a prediction activity and its accompanying prediction claim in mechanistic philosophy can be understood both ways. First, in the previous section, I have claimed that predictions of outcomes of medical interventions ought to be based on models of mechanisms, and therefore, a prediction claim about the outcome of intervention is a kind of causal explanation. On the other hand, I have presented ideas from the discussion in the mechanistic literature where a prediction claim is understood in a different way. There, prediction claims based on models of mechanisms serve as hypotheses about the outcomes of proposed models of mechanisms. That is, they are hypotheses for testing whether a proposed model's component parts, activities/interactions, and organization can together be causally able to produce a phenomenon. Therefore, when diagnosis is inferred from the knowledge of mechanisms (evidence of mechanisms), it amounts to a retrodictive claim – namely, *a hypothesis about the presence or occurrence of a particular type of mechanism in the body based on observable outcomes, outputs, or phenomena*. Thus:

> **MR6.** Mechanistic reasoning in diagnosis amounts to an inference (a hypothesis) about the presence of a type of mechanism causally or constitutively responsible for symptoms and signs.

If diagnosis is a product of mechanistic reasoning, then I take it that it is a retrodictive claim since, according to the discussion from the first section of this chapter, the thing it predicts – an instantiation or occurrence of a pathological mechanism (or dysfunctional physiological mechanism, depending on which kind of explanatory view of disease is assumed)

– is not a future event but an event or state of affairs that has already occurred or that is still persisting.

Thagard and Maung, for example, seem to hold the view asserting that the provision of a diagnosis seems to posit a story about how your symptoms and signs are caused by an underlying biological causal structure. Consider the following passage by Maung: "Where a diagnosis serves as an explanation of patient data, it does so partly by *denoting* a kind of causal structure that is instantiated by the actual patient. For example, the diagnosis of acute appendicitis *explains* a patient's abdominal pain by denoting a distinctive pathological type, in this case acute inflammation of the appendix, which is causing the abdominal pain" (Maung 2019: 513, emphases added). I will take the notions in italics from this passage at face value. Here, then, Maung assumes that diagnosis works as an explanation by proxy. The diagnostic claim stands in some sort of denoting relation to the causal structure that ought to explain the symptoms. But denoting is not explaining and pointing to possible explanations from a store of possible explanations is not itself part of an explanation. It can be part of a process of constructing an explanation, but it is not explanation itself. Benzi's view, for example, possibly points in the same direction: "once formulated, however, a diagnosis can be synthetically described, from a statistical viewpoint, as a relation between a set of findings (signs, symptoms, laboratory test results) and a certain pathological condition attributed to the patient" (Benzi 2013: 365). The diagnostic claim that a patient with a particular set of symptoms has acute appendicitis does not explain how acute appendicitis causes abdominal pain and fever (no matter whether it is taken as a causal explanation of a type of phenomenon – acute appendicitis – or a token phenomenon – patient *X*'s acute appendicitis). A diagnostic claim that a patient had a SARS-COV-2 infection is not an explanation of that patient's post-Covid symptoms, such as hair loss, or enduring fatigue. If we uphold the epistemic conception of explanation, as I did in Chapter II, then acute appendicitis by itself does not explain abdominal pain and fever. Diagnosis, therefore, does not explain symptoms and signs. Diagnosis serves as a hypothesis about the presence of a certain structure that can be, quite independently, considered causally responsible for symptoms and signs. But there is a big step from asserting a claim that this patient has acute appendicitis and the explanation that this particular acute appendicitis causes patient's abdominal pain or that in general acute appendicitis causes abdominal pain.

Once a diagnosis claim has been asserted, a similar algorithm to the one in Figure 10 ensues. Here, however, the hypothesis is not a claim about mechanism parts, their properties,

activities, interactions, or organizational features. Rather, the hypothesis is about the presence of a certain type of mechanism behind an observed phenomenon. Diagnostic tests are then performed, and doctors often intervene into some entity to see whether it will have an expected effect. If these tests, designed to reveal characteristics of certain types of mechanisms and clinically trialed for the assessment of their efficacy, reveal changes in the outcomes predicted by a diagnostic claim, then the mechanism underlying the symptoms and signs is confirmed at best, or corroborated at worst. However, contrary to the usual understanding of mechanistic reasoning, I have argued that it neither necessarily includes interventions into mechanism as a means of medical treatments nor that the prediction claim is this about the outcome of that intervention. I have also claimed that the type of inference used will depend on the particular model of a mechanism. Mechanistic reasoning, as I have proposed in statement MR1, refers to nothing over and above the prediction activity, of which the final product is a claim about the mechanism's output. So, diagnostic claims or hypotheses can be confirmed or corroborated by observational evidence or evidence from non-intervention. These examples are not obscure. In fact, they are regularly performed in clinical practice. For example, simple blood analyses to measure red blood cells or hemoglobin and magnetic resonance imaging do not seem to be interventions into mechanisms in any of the ways that, for example, Craver and Darden have mentioned and discussed (see section 3.4.).

To conclude, many diseases (especially mental diseases and cases such as rosacea) just seem to be collections of symptoms and by that, their diagnoses seem to be nothing over and above categorization. But the account presented here is not intended to be a full account of medical diagnosis. It is intended as an analysis of mechanistic reasoning in diagnosis. If there is no model of mechanism to reason upon, then there simply cannot be mechanistic reasoning. If, on the other hand, we reason upon a model of mechanism to assert a diagnostic claim, then, I argue, it is a hypothesis about the instantiation of a particular type of mechanism. It is not an explanation, but it can be constitutive of an explanation-making process.

CONCLUSION

In the early 1990s, new ideas appeared in medicine and philosophy of science. In medicine, the Evidence Based Medicine movement significantly influenced medical science, practice, and education. From its beginning, the movement aimed at changing clinical decision-making and clinical practice by increasing the use of evidence gathered by characteristic epidemiological studies (both experimental and observational). Its evidential framework was supposed to replace old-fashioned clinical decision-making which was based on expert knowledge, anecdotal evidence, and evidence gathered from the studies of medicine's laboratory sciences. Meanwhile, mechanistic philosophy was slowly gaining momentum in the 1990s, and by the early 2000s mechanisms and mechanistic explanations were among the most discussed notions in philosophy of science. Mechanistic philosophy embraced scientific realism and a bottom-up methodology - the practice of biological sciences was supposed to give answers to the philosophical questions of causality, scientific explanation, and prediction. Hence, mechanistic philosophers argued that the bulk of explanations in biological sciences looked like models of mechanisms rather than anything else. In years after they first appeared, both movements became immensely influential in their fields. Much of contemporary medicine is predominantly evidence-based in the way argued for by the proponents of the EBM movement. Similarly, many philosophers claim that by now there is a comprehensive philosophical framework with clearly discernible ontological and epistemological commitments – "The New Mechanistic Philosophy".

However, as I discussed throughout this dissertation, the main and most interesting point of difference between these two movements is in the kind of evidence each movement considers to be the best for the assessment of predictions of the outcomes of medical interventions. As I argued, the EBM's argument for the unreliability of mechanistic reasoning in the assessment of prediction claims is a conclusion of enumerative induction – mechanistic reasoning recurrently fails and therefore will continue to fail. Mechanistic philosophers, on the other hand, emphasize the knowledge of details regarding components of mechanisms in achieving explanatory goals but also in predicting interventions into mechanisms. Although mechanistic philosophers have taken into consideration this divergence of ideas before, this dissertation is a first comprehensive overview of the relation between mechanistic philosophy's core commitments and the contemporary EBM-influenced medical science and practice.

In each chapter of this dissertation, I discussed one specific question concerning mechanisms and the mechanistic approach in medicine. In the first chapter I defined the main characteristics of mechanistic and EBM-favored approaches to the investigation of disease causation and the methods and evidence needed to provide explanation and prediction. In the second and third chapter I provided my account of mechanistic explanation and prediction with a special focus on medical science and practice. Therefore, in addressing these specific questions I have shaped my own account of mechanistic explanation and prediction. Most importantly, it is an account grounded in the disambiguation between three theses of mechanistic philosophy, where each thesis corresponds to a set of ontological, epistemological, and methodological claims of mechanistic philosophy.

My account of mechanistic explanation and prediction can be characterized as a liberal view of the ideas of "The New Mechanistic Philosophy". This view intentionally neglected and diminished the role of ontological mechanisms and openly favored the role of epistemic features and characteristics of models of mechanisms in the construction of mechanistic explanations and predictions and in the assessment of their quality. Certainly, this generates other problems, some of which, I believe, have been successfully addressed. Nonetheless, I take that my liberal view of what counts as a mechanism and how to assess the quality of mechanistic explanation and prediction comes as a consequence of being honest to the bottom-up methodology embraced by mechanistic philosophers. Also, it provided resources for answering the question regarding the main point of divergence between mechanistic philosophy and the EBM movement stated above: why mechanistic reasoning fails so often and how we can make it better.

REFERENCES:

[1] Andersen, H. (2012). Mechanisms: what are they evidence for in evidence-based medicine?. *Journal of evaluation in clinical practice*, *18*(5), 992-999.

[2] Anić, Z. (2021). The Metaphysics of Causation in Biological Mechanisms: A Case of the Genetic Switch in Lambda Phage. *Acta Biotheoretica*, *69*(3), 435-448.

[3] Anscombe, G.E.M. (1973). Causality and Determination. In Sosa, E., & Tooley, M. (1993). *Causation*. Oxford: Oxford University Press (pp. 88-104)

[4] APS (2006). Strategic Plan 2006–2010. American Physiological Society, Bethesda, MD.

[5] Aronson, J. K. (2020). Defining aspects of mechanisms: evidence-based mechanism (evidence for a mechanism), mechanism-based evidence (evidence from a mechanism), and mechanistic reasoning. In *Uncertainty in Pharmacology* (pp. 3-38). Springer, Cham.

[6] Aronson, J. K., La Caze, A., Kelly, M. P., Parkkinen, V. P., & Williamson, J. (2018). The use of mechanistic evidence in drug approval. *Journal of Evaluation in Clinical Practice*, *24*(5), 1166-1176.

[7] Aronson, J. K., Auker-Howlett, D., Ghiara, V., Kelly, M. P., & Williamson, J. (2021). The use of mechanistic reasoning in assessing coronavirus interventions. *Journal of evaluation in clinical practice*, *27*(3), 684-693.

[8] Ashcroft, R. E. (2005). Current epistemological problems in evidence-based medicine. *Evidence-based Practice in Medicine and Health Care*, 77-85.

[9] Baetu, T. M. (2015). The completeness of mechanistic explanations. *Philosophy of Science*, *82*(5), 775-786.

[10] Barnes, E. C. (2005). Predictivism for pluralists. *The British journal for the philosophy of science*, *56*(3), 421-450.

[11] Barnes, E. C. (2008). *The paradox of predictivism*. Cambridge: Cambridge University Press.

[12] Barros, D. B. (2008). Natural selection as a mechanism. *Philosophy of Science*, *75*(3), 306-322.

[13] Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science*, *84*(2), 214-233.

[14] Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731-756.

[15] Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of science*, *78*(4), 533-557.

[16] Bechtel, W. (2019). Analysing network models to make discoveries about biological mechanisms. *The British Journal for the Philosophy of Science*, *70*(2), 459-484.

[17] Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421-441.

[18] Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT press.

[19] Bechtel, W., Abrahamsen, A., & Sheredos, B. (2018). Using diagrams to reason about biological mechanisms. In *International Conference on Theory and Application of Diagrams* (pp. 264-279). Springer, Cham.

[20] Benzi, M. (2011). Medical diagnosis and actual causation. *L&PS—Logic and Philosophy of Science*, *9*, 365-372.

[21] Bernard, C., Wolf, S., & Copley Greene, H. (1999). Experimental Medicine (1st ed.). Routledge.

[22] Bich, L., & Bechtel, W. (2021). Mechanism, autonomy and biological explanation. *Biology & Philosophy*, *36*(6), 1-27.

[23] Bluhm, R. (2013). Physiological mechanisms and epidemiological research. *Journal of Evaluation in Clinical Practice*, *19*(3), 422-426.

[24] Bluhm, R., & Borgerson, K. (2011). Evidence-based medicine. In Gabbay, D. M., Gifford, F., Thagard, P., & Woods, J. (Eds.). *Philosophy of medicine* (pp. 203-238). North-Holland.

[25] Bogen, J. (2008). Causally productive activities. *Studies in History and Philosophy of Science Part A*, *39*(1), 112-123.

[26] Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The philosophical review*, *97*(3), 303-352.

[27] Bolinska, A. (2013). Epistemic representation, informativeness and the aim of faithful representation. *Synthese*, *190*(2), 219-234.

[28] Boniolo, G., & Campaner, R. (2018). Molecular pathways and the contextual explanation of molecular functions. *Biology & Philosophy*, *33*(3), 1-19.

[29] Boorse, C. (1975). On the Distinction between Disease and Illness. *Philosophy & Public Affairs*, *5*(1), 49–68.

[30] Boorse, C. (1977). Health as a theoretical concept. *Philosophy of science*, *44*(4), 542-573.

[31] Boorse, C. (1997). A rebuttal on health. In Humber, J.M., Almeder, R.F. (Eds.), *What is disease?* (pp. 1-134). Humana Press, Totowa, NJ.

[32] Boorse, C. (2011). Concepts of health and disease. In *Philosophy of medicine* (pp. 13-64). North-Holland.

[33] Broadbent, A. (2009). Causation and models of disease in epidemiology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *40*(4), 302-311.

[34] Broadbent, A. (2011). Inferring causation in epidemiology: mechanisms, black boxes, and contrasts. In Illari, P. M., Russo, F., & Williamson, J. (Eds.). *Causality in the Sciences*. Oxford University Press (pp. 45-69).

[35] Broadbent, A. (2013). *Philosophy of Epidemiology*. Palgrave MacMillan.

[36] Brzović, Z., Balorda, V., & Šustar, P. (2021). Explanatory hierarchy of causal structures in molecular biology. *European Journal for Philosophy of Science*, *11*(2), 1-21.

[37] Capote, L., Nyakundi, R., Martinez, B., Lymperopoulos, A. (2015), Pathophysiology of Heart Failure. In Jagadeesh, G., Balakumar, P., & Maung-U, K. (Eds.). *Pathophysiology and pharmacotherapy of cardiovascular disease* (pp. 37-56). Springer International Publishing.

[38] Carrier, M. (2014). Prediction in context: On the comparative epistemic merit of predictive success. *Studies in History and Philosophy of Science Part A*, *45*, 97-102.

[39] Carter, K. C. (1987). Edwin Klebs' criteria for disease causality. *Medizinhistorisches Journal*, (H. 1), 80-89.

[40] Carter, K. C. (2017). *The rise of causal concepts of disease: case histories*. Routledge.

[41] Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University press.

[42] Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Clarendon Press.

[43] Cartwright, N. (2004). Causation: One word, many things. *Philosophy of Science, 71*(5), 805-819.

[44] Cartwright, Nancy. *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press, 2007.

[45] Chao, H. K., Chen, S. T., & Millstein, R. L. (2013) Towards the Methodological Turn in the Philosophy of Science. In Chao, H. K., Chen, S. T., & Millstein, R. L. (Eds.). *Mechanism and causality in biology and economics* (pp.1-18). Dordrecht: Springer.

[46] Chin-Yee, B. H. (2014). Underdetermination in evidence-based medicine. *Journal of evaluation in clinical practice*, *20*(6), 921-927.

[47] Clarke, B. (2011). *Causality in medicine with particular reference to the viral causation of cancers*. Dissertation, University College London.

[48] Clarke, B., & Russo, F. (2018). Mechanisms and biomedicine. In Glennan, S., Illari, P. (Eds.). *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 319-331). London, UK: Routledge.

[49] Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine*, *57*(6), 745-747.

[50] Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, *33*(2), 339-360.

[51] Claveau, F. (2012). The Russo-Williamson Theses in the social sciences: Causal inference drawing on two types of evidence. *Stud Hist Philos Biol Biomed Sci* 43 (4):806–13.

[52] Colombo, M., Hartmann, S., & van Iersel, R. (2014). Models, mechanisms, and coherence. The British Journal for the Philosophy of Science, 66(1), 181–212.

[53] Couch, M. B. (2011). Mechanisms and constitutive relevance. *Synthese*, *183*(3), 375-388.

[54] Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of science*, *68*(1), 53-74.

[55] Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355-376.

[56] Craver, C.F. (2007a). *Explaining the Brain*. Oxford: Oxford University Press.

[57] Craver, C.F. (2007b). Constitutive explanatory relevance. *Journal of Philosophical Research*, *32*, 3-20.

[58] Craver, C. F. (2013). Functions and mechanisms: A perspectivalist view. In Huneman, P. (Ed.), *Functions: Selection and mechanisms* (pp. 133-158). Springer, Dordrecht.

[59] Craver, C. F. (2014). The ontic account of scientific explanation. In Kaiser, M. I., Scholz, O. R., Plenge, D., & Hüttemann, A. (Eds), *Explanation in the special*

sciences. The Case of Biology and History, Dordrecht (pp. 27-52). Springer, Dordrecht.

[60] Craver, C.F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & philosophy*, *22*(4), 547-563.

[61] Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.

[62] Craver, C. F., & Kaplan, D. M. (2020). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, *71*(1), 287-319.

[63] Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, 1-22.

[64] Cummins, R. (1975). Functional analysis. *Journal of Philosophy*. 72 (November): 741-64.

[65] Cummins, R. (2000). How does it work?" versus" what are the laws?": Two conceptions of psychological explanation. *Explanation and cognition*, 117-144.

[66] Dammann, O. (2020). *Etiological Explanations: Illness Causation Theory*. CRC Press.

[67] Darden, L. (2006). *Reasoning in biological discoveries: Essays on mechanisms, interfield relations, and anomaly resolution*. Cambridge University Press.

[68] Darden, L. (2013). Mechanisms versus causes in biology and medicine. In Chao, H. K., Chen, S. T., & Millstein, R. L (Eds.), *Mechanism and causality in biology and economics* (pp. 19-34). Springer, Dordrecht.

[69] Darden, L. (2018). Strategies for discovering mechanisms. In Glennan, S., Illari, P. (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 255-266). Routledge.

[70] Darden, L., & Craver, C. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *33*(1), 1-28.

[71] Darden, L., Kundu, K., Pal, L. R., & Moult, J. (2018). Harnessing formal concepts of biological mechanism to analyze human disease. *PLOS Computational Biology*, *14*(12), e1006540.

[72] Davidoff, F., Haynes, B., Sackett, D., & Smith, R. (1995). Evidence based medicine. *Bmj*, *310*(6987), 1085-1086.

[73] Davies, P. S. (2000). Malfunctions. *Biology and Philosophy*, 15, 19–38.

[74] De Tejada, I. S., Angulo, J., Cellek, S., González-Cadavid, N., Heaton, J., Pickard, R., & Simonsen, U. (2005). Pathophysiology of erectile dysfunction. *The journal of sexual medicine*, *2*(1), 26-39.

[75] De Vreese, L., Weber, E., & Van Bouwel, J. (2010). Explanatory pluralism in the medical sciences: theory and practice. *Theoretical medicine and bioethics*, *31*(5), 371-390.

[76] Djulbegovic, B., Guyatt, G. H., & Ashcroft, R. E. (2009). Epistemologic inquiries in evidence-based medicine. *Cancer control*, *16*(2), 158-168.

[77] Doll, R., & Peto, R. (1976). Mortality in relation to smoking: 20 years' observations on male British doctors. *Br med J*, *2*(6051), 1525-1536.

[78] Douglas, H. E. (2009). Reintroducing prediction to explanation. *Philosophy of Science*, *76*(4), 444-463.

[79] Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of science*, *59*(2), 195-216.

[80] Duffy, T. P. (2011). The Flexner report—100 years later. *The Yale journal of biology and medicine*, *84*(3), 269.

[81] Eckardt, B. V., & Poland, J. S. (2004). Mechanism and explanation in cognitive neuroscience. *Philosophy of science*, *71*(5), 972-984.

[82] Eells, E. (1991). *Probabilistic causality* (Vol. 1). Cambridge University Press.

[83] Engelhardt, T. (1986) *The foundations of bioethics*, New York: Oxford University Press.

[84] Ereshefsky, M. (2009). Defining 'health' and 'disease'. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *40*(3), 221-227.

[85] Evans, A. S. (1993). *Causation and disease: a chronological journey*. Springer Science & Business Media.

[86] Fiorentino, A. R., & Dammann, O. (2015). Evidence, illness, and causation: An epidemiological perspective on the Russo–Williamson Thesis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *54*, 1-9.

[87] Fisher, R. A. (1958). Cancer and smoking. *Nature*, *182*(4635), 596-596.

[88] Frigg, Roman and Stephan Hartmann, "Models in Science", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/models-science/>.

[89] Fuller, J. (2016). *The new medical model: chronic disease and evidence-based medicine*. Doctoral dissertation. University of Toronto, Canada.

[90] Fuller, J. (2017). The new medical model: a renewed challenge for biomedicine. *CMAJ*, *189*(17), E640-E641.

[91] Fuller, J. (2021). The myth and fallacy of simple extrapolation in medicine. *Synthese*, *198*(4), 2919-2939.

[92] Fuller, J., & Flores, L. J. (2015). The Risk GP Model: The standard model of prediction in medicine. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *54*, 49-61.

[93] Fuller, J., & Flores, L. J. (2016). Translating trial results in clinical practice: the risk GP model. *Journal of cardiovascular translational research*, *9*(3), 167-168.

[94] Garson, J. (2013). The functional sense of mechanism. *Philosophy of science*, *80*(3), 317-333.

[95] Garson, J. (2016). *A critical overview of biological functions*. Cham: Springer International Publishing.

[96] Garson, J. (2018). Mechanisms, phenomena, and functions. In Glennan, S., Illari, P. (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 104-115). Routledge.

[97] Gebharter, A. (2017). Uncovering constitutive relevance relations in mechanisms. *Philosophical Studies*, *174*(11), 2645-2666.

[98] Ghofrani, H. A., Osterloh, I. H., & Grimminger, F. (2006). Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. *Nature reviews Drug discovery*, *5*(8), 689-702.

[99] Giere, R. N. (1999). Using models to represent reality. In L. Magnani, N. Nerssesian, & P. Thagard (Eds.), Model-based reasoning in scientific discovery (pp. 41–57). New York: Kluwer Academic.

[100] Giere, R. N. (2004). How models are used to represent reality. Philosophy of Science, 71(5), 742–752.

[101] Gillies, D. (2011). The Russo-Williamson thesis and the question of whether smoking causes heart disease. In Illari, P. M., Russo, F., & Williamson, J. (Eds.). *Causality in the Sciences*. Oxford University Press. (pp. 110-125).

[102] Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*(1), 49-71.

[103] Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of science*, *69*(S3), S342-S353.

[104] Glennan, S. (2009). Productivity, relevance and natural selection. *Biology & Philosophy*, *24*(3), 325-339.

[105] Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis*, *72*(2), 251-266.

[106] Glennan, S. (2011). Singular and general causal relations: A mechanist perspective. In Illari, P. M., Russo, F., & Williamson, J. (Eds.), *Causality in the Sciences* (pp. 789) Oxford University Press.

[107] Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.

[108] Glennan, S., Illari, P. (2018). Varieties of mechanisms. In Glennan, S., Illari, P. (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 91-103). Routledge.

[109] Greenhalgh, T. (2019). *How to read a paper: the basics of evidence-based medicine*. John Wiley & Sons.

[110] Glymour, M. M., & Spiegelman, D. (2017). Evaluating public health interventions: 5. Causal inference in public health research—do sex, race, and biological factors cause health outcomes?. *American journal of public health*, *107*(1), 81-85.

[111] Godfrey-Smith, P. (2009). Abstractions, Idealizations, and Evolutionary Biology. In A. Barberousse, M. Morange, & T. Pradeu (Eds.), Mapping the future of biology: Evolving concepts and theories. Berlin: Springer.

[112] Guyatt, G., Cairns, J., Churchill, D., Cook, D., Haynes, B., Hirsh, J., ... & Tugwell, P. (1992). Evidence-based medicine: a new approach to teaching the practice of medicine. *Jama*, *268*(17), 2420-2425.

[113] Guyatt, G., Rennie, D., Meade, M., & Cook, D. (Eds.). (2015). *Users' guides to the medical literature: a manual for evidence-based clinical practice, Third Edition*, McGraw Hill/Medical.

[114] Hafeman, D. M., & Schwartz, S. (2009). Opening the Black Box: a motivation for the assessment of mediation. *International journal of epidemiology*, *38*(3), 838-845.

[115] Hall, N. (2004). Two concepts of causation. In Collins, J., Hall, N., & Paul, L. A. (eds) *Causation and counterfactuals*. 225-276. Mit Press.

[116] Haufe, C. (2013). From necessary chances to biological laws. *The British journal for the philosophy of science*, *64*(2), 279-295.

[117] Haynes, B., Sacket, D.L., Guyatt, G.H., Tugwell, P. (2005). *Clinical Epidemiology: How to do clinical practice research*. Lippincott Williams & Wilkins.

[118] Hempel, C. G. (1966). *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.

[119] Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of science*, *15*(2), 135-175.

[120] Hennekens, C.H., Buring, J.E. (1987). *Epidemiology in medicine*. Boston: Litlle, Brown.

[121] Hernán, M.A., Robins, J.M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

[122] Hesslow, G. (1993). Do we need a concept of disease?. *Theoretical medicine*, *14*(1), 1-14.

[123] Hill, A. B. (1965). The environment and disease: association or causation?. *Proceedings of the Royal Society of Medicine*, *58*(5), 295–300.

[124] Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945-960.

[125] Howick, J. (2011a). *The philosophy of evidence-based medicine*. John Wiley & Sons.

[126] Howick, J. (2011b). Exposing the vanities—and a qualified defense—of mechanistic reasoning in health care decision making. *Philosophy of Science*, *78*(5), 926-940.

[127] Howick J, Glasziou P, Aronson J.K. Evidence-based mechanistic reasoning. *J R Soc Med*. 2010;103(11):433-441.

[128] Howick, J., Glasziou, P., & Aronson, J.K. (2013). Problems with using mechanisms to solve the problem of extrapolation. *Theoretical medicine and bioethics*, *34*(4), 275-291.

[129] Howson, C., Urbach P.M. (1993). *Scientific Reasoning - the Bayesian Approach*, 2nd ed. Chicago and La Salle: Open Court.

[130] Hu, F. (2008). *Obesity epidemiology*. Oxford University Press.

[131] Illari, P. M. (2011). Mechanistic evidence: disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science*, *25*(2), 139-157.

[132] Illari, P. (2013). Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, *78*(2), 237-255.

[133] Illari, P. M., & Williamson, J. (2011). Mechanisms are real and local. In Illari, P. M., Russo, F., & Williamson, J. (Eds.). *Causality in the Sciences* (pp. 818-844) Oxford University Press.

[134] 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.

[135] Institute of Medicine (US) Committee to Advise the Public Health Service on Clinical Practice Guidelines, Field, M. J., & Lohr, K. N. (Eds.). (1990). *Clinical Practice Guidelines: Directions for a New Program*. National Academies Press (US).

[136] Ioannidis, S., & Psillos, S. (2017). In defense of methodological mechanism: The case of apoptosis. *Axiomathes*, *27*(6), 601-619.

[137] Ioannidis, S., & Psillos, S. (2018). Mechanisms in practice: A methodological approach. *Journal of evaluation in clinical practice*, *24*(5), 1177-1183.

[138] Jarada, T. N., Rokne, J. G., & Alhajj, R. (2020). A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *Journal of cheminformatics*, *12*(1), 1-23.

[139] Jerkert, J. (2015). Negative mechanistic reasoning in medical intervention assessment. *Theoretical medicine and bioethics*, *36*(6), 425-437.

[140] Jones, M. (2005). Idealization and abstraction: A framework. In M. R. Jones & N. Cartwright (Eds.), Idealization XII: Correcting the model. Idealization and abstraction in the sciences (Vol. 86, pp. 173–217). Amsterdam: Rodopi.

[141] Kaiser, M.I. (2018). The components and boundaries of mechanisms. In Glennan, S., Illari, P. (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 116-130). Routledge.

[142] Kannel, W. B., Skinner, J. J., Jr, Schwartz, M. J., & Shurtleff, D. (1970). Intermittent claudication. Incidence in the Framingham Study. *Circulation*, *41*(5), 875–883.

[143] Kannel, W. B., Castelli, W. P., Gordon, T., & McNamara, P. M. (1971). Serum cholesterol, lipoproteins, and the risk of coronary heart disease. The Framingham study. *Annals of internal medicine*, *74*(1), 1–12.

[144] Kannel, W. B., Castelli, W. P., McNamara, P. M., McKee, P. A., & Feinleib, M. (1972). Role of blood pressure in the development of congestive heart failure: the Framingham study. *New England Journal of Medicine*, *287*(16), 781-787.

[145] Kannel, W. B., McGee, D., & Gordon, T. (1976). A general cardiovascular risk profile: the Framingham Study. *The American journal of cardiology*, *38*(1), 46-51.

[146] Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339-373.

[147] 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.

[148] Kästner, L., & Andersen, L. M. (2018). Intervening into mechanisms: Prospects and challenges. *Philosophy Compass*, *13*(11), e12546.

[149] Kauffman S.A. (1971) Articulation of Parts Explanation in Biology and the Rational Search for them. In: Buck R.C., Cohen R.S. (eds) PSA 1970. Boston Studies in the Philosophy of Science, vol 8. Springer, Dordrecht.

[150] Kim, J. (1964). Inference, explanation, and prediction. *The journal of philosophy*, *61*(12), 360-368.

[151] Kincaid, H. (2011). Causal modelling, mechanism, and probability in epidemiology. In Illari, P. M., Russo, F., & Williamson, J. (Eds.). *Causality in the Sciences* (pp. 70-90). Oxford University Press.

[152] Kingma, E. (2007). What is it to be healthy?. *Analysis*, *67*(2), 128-133.

[153] Kitcher, P. (1984). 1953 and all that. A tale of two sciences. *The Philosophical Review*, *93*(3), 335-373.

[154] Kots, A. Y., Martin, E., Sharina, I. G., & Murad, F. (2009). A short history of cGMP, guanylyl cyclases, and cGMP-dependent protein kinases. *cGMP: Generators, Effectors and Therapeutic Implications*, 1-14.

[155] Kohár, M., & Krickel, B. (2021). Compare and Contrast: How to Assess the Completeness of Mechanistic Explanation. In Calzavarini, F., & Viola, M. (Eds.), *Neural Mechanisms: New Challenges in the Philosophy of Neuroscience* (Vol. 17) (pp. 395-424. Springer Nature.

[156] Krieger, N. (1994). Epidemiology and the web of causation: has anyone seen the spider?. *Social science & medicine*, *39*(7), 887-903.

[157] Krieger, N. (2011). *Epidemiology and the people's health: theory and context*. Oxford University Press.

[158] La Caze, A. (2011). The role of basic science in evidence-based medicine. *Biology & Philosophy*, *26*(1), 81-98.

[159] Lakhani, S. R., Finlayson, C. J., Dilly, S. A., & Gandhi, M. (2016). *Basic pathology: An introduction to the mechanisms of disease*. Crc Press.

[160] Lancaster Jr, J. R. (2017). A Concise History of the Discovery of Mammalian Nitric Oxide (Nitrogen Monoxide) Biogenesis. In *Nitric Oxide* (pp. 1-7). Academic Press.

[161] Lange, M. (2016). *Because Without Cause: Non-Casual Explanations In Science and Mathematics*. Oxford University Press.

[162] Lappi, O., & Rusanen, A. M. (2011). Turing machines and causal mechanisms in cognitive science. In Illari, P. M., Russo, F., & Williamson, J. (Eds.). *Causality in the Sciences* (pp. 224-239). Oxford University Press.

[163] Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *The British Journal for the Philosophy of Science, 63*(2), 399-427.

[164] Levy, A. (2013). Three kinds of new mechanism. *Biology & Philosophy, 28*(1), 99-114.

[165] Levy, A. (2021). Idealization and abstraction: refining the distinction. *Synthese, 198*(24), 5855-5872.

[166] Levy, A., & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of science, 80*(2), 241-261.

[167] Lewis, D. (1973). Causation. *The journal of philosophy,* 70(17), 556-567.

[168] Lipton, P. (1990). Prediction and prejudice. *International Studies in the Philosophy of Science, 4*(1), 51-65.

[169] Lipton, P. (1993). Making a Difference. *Philosophica* 51: 39-54.

[170] Lipton, P. (2003). *Inference to the best explanation*. Routledge.

[171] Lipton, R., & Ødegaard, T. (2005). Causal thinking and causal language in epidemiology: it's in the details. *Epidemiologic Perspectives & Innovations*, *2*(1), 1-9.

[172] Love, A. C., & Nathan, M. J. (2015). The idealization of causation in mechanistic explanation. *Philosophy of Science, 82*(5), 761-774.

[173] Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International studies in the philosophy of science*, *18*(1), 27-39.

[174] Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science*, *67*(1), 1-25.

[175] Mackie, J. L. (1965). Causes and conditions. *American philosophical quarterly*, *2*(4), 245-264.

[176] Mackie, J. L. (1980). *The cement of the universe: A study of causation*. Clarendon Press.

[177] MacMahon, B., & Pugh, T. F. (1970). Epidemiology: principles and methods. *Epidemiology: principles and methods.*

[178] Magnani, L., & Bertolotti, T. (Eds.). (2017). *Springer handbook of model-based science*. Springer.

[179] Marcum, J.A. (2008). *An introductory philosophy of medicine: Humanizing modern medicine*. Vol. 99. Springer Science & Business Media.

[180] Maung, H. H. (2019). The Functions of Diagnoses in Medicine and Psychiatry. In Tekin, S., & Bluhm, R. (Eds.). (2019). *The Bloomsbury companion to philosophy of psychiatry* (pp. 507-526). Bloomsbury Publishing.

[181] Mavromoustakos, T., Durdagi, S., Koukoulitsa, C., Simcic, M., G Papadopoulos, M., Hodoscek, M., & Golic Grdadolnik, S. (2011). Strategies in the rational drug design. *Current medicinal chemistry*, *18*(17), 2517-2530.

[182] Mill, J. S. 1858 (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Princilples of Evidence and the Methods of Scientific Investigation*. Harper and brothers.

[183] Mills, K. T., Bundy, J. D., Kelly, T. N., Reed, J. E., Kearney, P. M., Reynolds, K., ... & He, J. (2016). Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation*, *134*(6), 441-450.

[184] Moghaddam-Taaheri, S. (2011). Understanding pathology in the context of physiological mechanisms: The practicality of a broken-normal view. *Biology & Philosophy*, *26*(4), 603-611.

[185] Mongtomery, K. (2005). *How doctors think: Clinical judgment and the practice of medicine*. Oxford University Press.

[186] Montori, V. M., & Guyatt, G. H. (2008). Progress in evidence-based medicine. *Jama*, *300*(15), 1814-1816.

[187] Morabia, A. (2006). Pierre-Charles-Alexandre Louis and the evaluation of bloodletting. *Journal of the Royal Society of Medicine*, *99*(3), 158-160.

[188] Moreno, A., Mossio, M. (2015). *Biological Autonomy: A Philosophical and Theoretical Inquiry*. Dordrecht: Springer.

[189] Moss, L. (2012). Is the philosophy of mechanism philosophy enough?. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 164-172.

[190] Mossio, M., Saborido, C., Moreno, A. (2009). An organizational account for biological functions. *The British Journal for the Philosophy of Science*, 60: 813–41.

[191] Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford University Press.

[192] Nathan, M. J. (2020). Causation by concentration. *The British Journal for the Philosophy of Science*, 65 (2):191-212.

[193] Nervi, M. (2010). Mechanisms, malfunctions and explanation in medicine. *Biology & Philosophy*, *25*(2), 215-228.

[194] Newton, W. (2001). Rationalism and empiricism in modern medicine. *Law and contemporary problems*, *64*(4), 299-316.

[195] 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.

[196] OCEBM Levels of Evidence Working Group, 2011. The Oxford 2011 Levels of Evidence. Oxford Centre for Evidence-Based Medicine. https://www.cebm.net/index.aspx?o=5653.

[197] Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., & Adhya, S. (2005). Switches in bacteriophage lambda development. *Annu. Rev. Genet.*, *39*, 409-429.

[198] Osterloh, I. H. (2004). The discovery and development of Viagra®(sildenafil citrate). In *Sildenafil* (pp. 1-13). Birkhäuser, Basel.

[199] Parascandola, M., & Weed, D. L. (2001). Causation in epidemiology. *Journal of Epidemiology & Community Health*, *55*(12), 905-912.

[200] Park K. (2019). A review of computational drug repurposing. *Translational and clinical pharmacology*, *27*(2), 59–63.

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

[202] Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.

[203] Pelling, M. (1997). Contagion/germ theory I specificity. In Bynum, W.F., Porter, R. (Eds). *Companion encyclopedia of the history of medicine* (pp. 309-334). Routledge.

[204] Peto R. (1984). The need for ignorance in cancer research. In Duncan R, Weston-Smith M (Eds.). *The Encyclopedia of Medical Ignorance*. Oxford, England: Pergamon Press. (pp. 129-133).

[205] Piccinini, G. (2007). Computing mechanisms. *Philosophy of Science*, *74*(4), 501-526.

[206] Portides, D. (2021). Idealization and abstraction in scientific modeling. *Synthese*, *198*(24), 5873-5895.

[207] Povich, M. (2021). Information and explanation: an inconsistent triad and solution. *European Journal for Philosophy of Science*, *11*(2), 1-17.

[208] Povich, M., & Craver, C. F. (2017). Mechanistic levels, reduction, and emergence. In Glennan, S., Illari, P. (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 185-197). Routledge.

[209] 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.

[210] Psillos, S. (2005). *Scientific realism: How science tracks truth*. Routledge.

[211] Ptashne, M. (2004). *A Genetic Switch, Third Edition: Phage Lambda Revisited* Cold Spring Harbor Laboratory Press.

[212] Putnam, H. (1975). *Philosophical Papers: Mathematics, matter, and method* (Vol. 1). CUP Archive.

[213] Reichenbach, H. (1956). *The direction of time* (Vol. 65). Univ of California Press.

[214] Reiss, J. (2012). Causation in the sciences: An inferentialist account. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(4), 769-777.

[215] Reiss, J. (2015). *Causation, evidence, and inference*. Routledge.

[216] Reiss, J., Ankeny, R.A. Philosophy of Medicine, *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/spr2022/entries/medicine/>.

[217] Reutlinger, A., & Saatsi, J. (Eds.). (2018). *Explanation beyond causation: philosophical perspectives on non-causal explanations*. Oxford University Press.

[218] Rizzi, D. A., & Pedersen, S. A. (1992). Causality in medicine: towards a theory and terminology. *Theoretical Medicine*, *13*(3), 233-254.

[219] Rose, G. (1992). *The Strategy of Preventive Medicine*. Oxford: Oxford University Press.

[220] Rose, G. (2001). Sick individuals and sick populations. *International journal of epidemiology*, *30*(3), 427-432.

[221] Ross, L. N. (2020). Causal concepts in biology: How pathways differ from mechanisms and why it matters. *The British Journal for the Philosophy of Science*, 72 (1), 131-158.

[222] Ross, L. N., & Woodward, J. F. (2016). Koch's postulates: an interventionist perspective. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *59*, 35-46.

[223] Rothman K. J. (1976). Causes. *American journal of epidemiology*, *104*(6), 587–592.

[224] Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (Vol. 3). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

[225] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

[226] Rusanen, A. M., & Lappi, O. (2007). The limits of mechanistic explanation in neurocognitive sciences. In *Proceedings of the European cognitive science conference* (pp. 284-289).

[227] Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International studies in the philosophy of science*, *21*(2), 157-170.

[228] Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *Bmj*, *312*(7023), 71-72.

[229] Sackett, D.L. (2000). The Fall of 'Clinical Research' and the Rise of 'Clinical-Practice Research'. Clinical and Investigative Medicine 23:379–81.

[230] Salmon, W. C. (1977). An" at-at" theory of causal influence. *Philosophy of Science*, *44*(2), 215-224.

[231] Salmon, W. C. (1984a). *Scientific explanation and the causal structure of the world*. Princeton University Press.

[232] Salmon, W. C. (1984b). Scientific explanation: Three basic conceptions. In *PSA: Proceedings of the biennial meeting of the philosophy of science association* (Vol. 1984, No. 2, pp. 293-305). Philosophy of Science Association.

[233] Salmon, W. C. (1998). *Causality and explanation*. Oxford University Press.

[234] Savitz D.A. (1994). In defense of black box epidemiology. Epidemiol. 5: 550-552.

[235] Schwartz, A., & Elstein, A. S. (2008). Clinical reasoning in medicine. *Clinical reasoning in the health professions*, *3*, 223-234.

[236] Shapiro, H. (2003). How different are Western and Chinese medicine? The case of nerves. In Selin, H. (ed.). *Medicine Across Cultures* (pp. 351-372). Springer, Dordrecht.

[237] Sheredos, B., Burnston, D., Abrahamsen, A., & Bechtel, W. (2013). Why do biologists use so many diagrams?. *Philosophy of Science*, *80*(5), 931-944.

[238] Simon, H. A. (1991). The architecture of complexity. *Proceedings of the American Philosophical Society* Vol. 106, No. 6 (Dec. 12, 1962), pp. 467-482.

[239] Simon, J. R. (2008). Constructive realism and medicine: An approach to medical ontology. *Perspectives in Biology and Medicine*, *51*(3), 353-366.

[240] Skipper Jr, R. A., & Millstein, R. L. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 327-347.

[241] Skrabanek, P. (1994). The emptiness of the black box. *Epidemiology*, 553-555.

[242] Smith, G. C., & Pell, J. P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *Bmj*, *327*(7429), 1459-1461.

[243] Solomon, M. (2011). Just a paradigm: evidence-based medicine in epistemological context. *European Journal for Philosophy of Science*, *1*(3), 451.

[244] Solomon, M. (2015). *Making medical knowledge*. Oxford University Press, USA.

[245] Solomon, M. (2016). On ways of knowing in medicine. *CMAJ*, *188*(4), 289-290.

[246] Spirtes, P., Glymour, C. N., Scheines, R. (1993). *Causation, prediction, and search*. Springer.

[247] Stanley, D. E., & Campos, D. G. (2013). The logic of medical diagnosis. *Perspectives in Biology and Medicine*, *56*(2), 300-315.

[248] Straus, S.E., Richardson, W.S., Glasziou, P., and Haynes, R.B. (2005). *Evidence-Based Medicine: How to Practice and Teach* EBM. Toronto: Elsevier.

[249] Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford University Press.

[250] Stempsey, W. E. (2000). A pathological view of disease. *Theoretical Medicine and Bioethics*, *21*(4), 321-330.

[251] Strevens, M. (2011). *Depth: An account of scientific explanation*. Harvard University Press.

[252] Suárez, M. (2003). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science*, 17, 225–244.

[253] Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of science*, *71*(5), 767-779.

[254] Susser, M. (1991). What is a cause and how do we know one? A grammar for pragmatic epidemiology. *American Journal of Epidemiology*, *133*(7), 635-648.

[255] Tabery, J. G. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of science*, *71*(1), 1-15.

[256] Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis (1975-)*, *55*(3), 393-415.

[257] Thagard, P. (1999). *How scientists explain disease*. Princeton University Press.

[258] Thagard, P. (2003). Pathways to biomedical discovery. *Philosophy of science*, *70*(2), 235-254.

[259] Thagard, P. (2005). 4 What is a medical theory?. *Studies in Multidisciplinarity*, *3*, 47–62.

[260] Tulodziecki, D. (2013). Shattering the myth of Semmelweis. *Philosophy of Science*, *80*(5), 1065-1075.

[261] Valles, S. (2020). Philosophy of Biomedicine. *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2020/entries/biomedicine/.

[262] van Eck, D. (2015). Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Axiomathes*, *25*(1), 5-22.

[263] Van Eck, D., & Gervais, R. (2016). Difference making, explanatory relevance, and mechanistic models. *THEORIA. Revista de Teoría, Historia y Fundamentos de la Ciencia*, *31*(1), 125-134.

[264] Vandenbroucke, J. P. (1988). Is 'the causes of cancer' a miasma theory for the end of the twentieth century?. *International journal of epidemiology*, *17*(4), 708-709.

[265] Vineis, P. (2004). Causality in epidemiology. In Morabia, A. (Ed.). *A history of epidemiologic methods and concepts*. Birkhäuser (pp. 337-351).

[266] Vohradsky J. (2017). Lambda phage genetic switch as a system with critical behaviour. *Journal of theoretical biology*, *431*, 32–38.

[267] Wakefield, J. (1992) "The concept of mental disorder: on the boundary between biological facts and social values, *American Psychologist*, 47, 373-388.

[268] Weber, M. (2004). *Philosophy of experimental biology*. Cambridge university press.

[269] Weber, M. (2009). The Crux of Crucial Experiments: Duhem's Problems and Inference to the Best Explanation. *The British Journal for the Philosophy of Science*, 60: 19-49.

[270] Weed, D. L. (1998). Beyond black box epidemiology. *American Journal of Public Health*, *88*(1), 12-14.

[271] Weed, D. L. (2000). Epidemiologic evidence and causal inference. *Hematology/oncology clinics of North America*, *14*(4), 797-807.

[272] Weed, D. L., & Hursting, S. D. (1998). Biologic plausibility in causal inference: current method and practice. *American Journal of Epidemiology*, *147*(5), 415-425.

[273] Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese*, *183*(3), 313-338.

[274] Werkhoven, S. (2019). A Dispositional Theory of Health. *The British Journal for the Philosophy of Science*, 70(4): 927-952.

[275] Whitbeck, C. (1981). What is diagnosis? Some critical reflections. *Metamedicine*, *2*(3), 319-329.

[276] Williamson, J. (2011). Mechanistic theories of causality part I. *Philosophy Compass*, *6*(6), 421-432.

[277] Williamson, J. (2019). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, *32*(1), 33-61.

[278] Wimsatt, W. C. (1972, January). Complexity and organization. In *PSA: Proceedings of the biennial meeting of the Philosophy of Science Association* (Vol. 1972, pp. 67-86). D. Reidel Publishing.

[279] Wimsatt, W. C. (1976). Reductive explanation: a functional account. In A. C. Michalos, C. A. Hooker, G. Pearce, and R. S. Cohen, eds., PSA-1974 (*Boston Studies in the Philosophy of Science*, volume 30, pp. 671–710). Dordrecht: Reidel.

[280] Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science*, *64*, S372-S384.

[281] Wimsatt, W.C. (2018). *Foreword*. In Glennan, S., Illari, P., *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 319-331). London, UK: Routledge.

[282] Woodward, J. (2002). What Is a Mechanism? A Counterfactual Account. *Philosophy of Science*, *69*(S3), S366-S377.

[283] Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford university press.

[284] Woodward, J. (2011). Mechanisms revisited. *Synthese*, *183*(3), 409-427.

[285] Woodward, J. (2013, June). II—Mechanistic explanation: Its scope and limits. In *Aristotelian Society Supplementary Volume* (Vol. 87, No. 1, pp. 39-65). Oxford, UK: Blackwell Publishing Ltd.

[286] Worrall, J. (2002). What evidence in evidence-based medicine?. *Philosophy of science*, *69*(S3), S316-S330.

[287] Worrall, J. (2014). Prediction and accommodation revisited. *Studies in History and Philosophy of Science Part A, 45*, 54-61.

[288] World Health Organization. (2014). Global status report on noncommunicable diseases 2014. World Health Organization.

[289] World Health Organization. (2018). Noncommunicable diseases country profiles 2018. World Health Organization.

[290] Wright, C. D. (2012). Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science*, *2*(3), 375-394.

[291] Wright, C., & Bechtel, W. (2006). Mechanisms and psychological explanation. In Thagard, P. (Ed.) *Philosophy of psychology and cognitive science* (pp. 31-79). North-Holland.

[292] Wright, C., & Van Eck, D. (2018). Ontic explanation is either ontic or explanatory, but not both. *ERGO-AN OPEN ACCESS JOURNAL OF PHILOSOPHY*, *5*(38), 997-1029.

[293] Yeh, R. W., Valsdottir, L. R., Yeh, M. W., Shen, C., Kramer, D. B., Strom, J. B., Secemsky, E.A., Healy, J.L., Domeier, R.M., Kazi, D.S., Nallamothu, B.K. (2018). Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *BMJ, 363*.

[294] Zernicke, R. F., & Whiting, W. C. (2000). Mechanisms of musculoskeletal injury. *Biomechanics in Sport. Oxford: Blackwell Science Ltd*, 507-22.

LIST OF FIGURES: