For forthcoming anthology, "Building Theories,"
Edited by Emiliano Ippoliti and David Danks, to be published by Springer,
based on workshop "Building Theories"
at Sapienza University, Rome, Italy, June 16-18, 2016

The Product Guides the Process: Discovering Disease Mechanisms

Lindley Darden, Lipika R. Pal, Kunal Kundu, and John Moult

Abstract

The nature of the product to be discovered guides the reasoning to discover it. Biologists and medical researchers often search for mechanisms. The "new mechanistic philosophy of science" provides resources about the nature of biological mechanisms that aid the discovery of mechanisms. Here, we apply these resources to the discovery of mechanisms in medicine. A new diagrammatic representation of a disease mechanism chain indicates both what is known and, most significantly, what is not known at a given time, thereby guiding the researcher and collaborators in discovery. Mechanisms of genetic diseases provide the examples.

Keywords: Discovery, Mechanism, Diagrams, Genetic Disease

Introduction

While physicists often represent theories as sets of mathematical laws, biologists usually represent general knowledge with schematic representations of mechanisms. Biologists and medical researchers seek mechanisms because knowing a mechanism facilitates explanation, prediction, and control. A theme in work on discovering mechanisms is captured by the slogan--the product guides the process. The thesis is that

characterizing a mechanism (the product) provides resources to guide the reasoning in its discovery (the process) (Craver and Darden 2013).

Recent philosophical analysis of mechanisms provides resources to aid the discovery of mechanisms. This work is being applied to the discovery of mechanisms in medicine. When the goal is to discover a disease mechanism, the nature of the product--the kind of disease mechanism--guides the process of searching for it. The kinds of products to be discussed here are representations of genetic disease mechanisms. In such diseases, genetic variants play a major role, together with environmental effects. The process is the reasoning to discover such mechanisms.

We develop a new graphical interface to aid medical researchers in hypothesizing and representing genetic disease mechanisms. We illustrate its use here in detailed diagrams of genetic mechanism schemas. The three examples are for a monogenic disease chain (cystic fibrosis), a cancer disease chain (affecting DNA mismatch repair), and one complex trait disease chain (one chain for one of many mutations for one of the loci associated with Crohn's disease).

This paper first summarizes recent work in mechanistic philosophy of science (e.g., Machamer, Darden, Craver 2000; Bechtel and Abrahamsen 2005; Glennan and Illari, forthcoming). As philosophers have shown, diagrams of mechanism schemas play important roles in abstractly representing the product to be discovered and guiding the process of discovery (e.g., Craver and Darden 2013; Abrahamsen and Bechtel 2015). A key idea is to sketch both what is known and what is not known at a given time. Black boxes in the sketch indicate where to fill in missing mechanism components. The next section of this paper reviews the application of the mechanistic perspective in the

philosophy of medicine (e.g., Thagard 1998; 1999; 2003; Darden 2013; Plutynski 2013). Then, we show how abstract mechanism chain diagrams serve to represent what is known or not known in genetic disease mechanisms. By depicting the state of knowledge about the genetic mechanism at a given time, the diagram perspicuously represents gaps in knowledge, namely, the sites of ignorance that researchers seek to remove. A set of heuristic questions provides guidance in filling the gaps. Three example diagrams of disease mechanism chains illustrate our new framework. We contrast our framework with two other graphical representation schemes. Finally, we propose future work, including plans for a web-based, graphical system that facilitates easy drawing and sharing of the individual mechanism chains, as well as discovery of interactions among them.

Mechanistic Philosophy of Science

Philosophers have been working for over twenty years to develop what is called the "new mechanistic philosophy of science" (Bechtel and Richardson 1993; Glennan 1996; Machamer, Darden, Craver 2000). This work calls attention to the importance of the search for mechanisms in biology and other disciplines, characterizes the nature of mechanisms, and compiles hindsight about the reasoning strategies used in the discovery of mechanisms (summarized in Craver and Darden 2013).

The discovery of a mechanism typically begins with a puzzling phenomenon. When the goal is to find what produces the phenomenon, then one searches for a mechanism. That decision rules out other parts of a large search space. One is not seeking merely a set of correlated variables. One is not seeking an economical equation that describes the phenomenon, although such an equation can provide a constraint in the search for a mechanism (Craver 2008; Bechtel and Abrahamsen 2013). One is not

seeking a law from which a description of the phenomenon can be derived. One is not merely seeking a relation between one cause and the phenomenon as the effect, although such a relation provides clues about mechanism components (Darden 2013). Nor is one merely seeking to find a pathway, characterized by nodes and unlabeled links which do not depict the activities that drive the mechanism. Rather, one is attempting to construct a mechanism schema that describes how entities and activities are spatially and temporally organized together to produce the phenomenon.

Employing a specific characterization of a mechanism provides guidance in discovery. One oft-cited mechanism characterization is this: "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, Craver 2000, p. 3). The goal in mechanism discovery is to find the entities and activities, to describe how they are organized, and to show how that productively continuous organization produces the phenomenon of interest. This characterization directs one to ask: What are the set up and finish conditions? Is there a specific, triggering start condition? What is spatially next to what? What is the temporal order of the steps? What are the entities in the mechanism? What are their structures? What are the activities that drive the mechanism? What are their range and their rate? How does each step of the mechanism give rise to the next? What are the activity enabling properties that make possible the next step? What are the activity signatures (properties of an entity or group of entities in a subsequent step) that show the kinds of activities that operated in the previous step to produce them? How was each step driven by the previous one? What is the overall organization of the mechanism: does it proceed linearly or is the mechanism perhaps cyclic (with no clear start and stop),

or is it organized with feedback loops, or does it have some other overall organizational motif? Where is it spatially located? In what context does the mechanism operate and how is it integrated with other mechanisms? These kinds of questions show how the nature of the product provides desiderata that guide the process of its discovery.

Mechanism schemas are representations of mechanisms. A "schema" (sometimes called a "model" of a mechanism) abstractly represents the structure of a target mechanism. Here is an example of a very abstract schema for the mechanism of protein synthesis: DNA→RNA→protein. Such schemas are often depicted in diagrams. William Bechtel and his collaborators (Sheredos et al. 2013; Abrahamsen and Bechtel 2015; Abrahamsen et al. forthcoming) discuss the many "visual heuristics" that diagrammatic representations of mechanism enable. They envisage biologists as reverse engineers, trying out various designs to spatially represent the interacting components of the mechanisms being discovered. The diagrams make salient specific aspects of the organization and operation of the mechanisms.

Schemas vary from one another along several dimensions: sketchy to sufficiently complete, abstract to specific, small to general scope of applicability, and possible to actual (Craver and Darden 2013, Ch. 3). A goal in discovering a mechanism is to convert an incomplete sketchy representation into an adequate one for the purpose at hand. Incomplete sketches indicate where black (unknown components) and grey (only functionally specified) boxes need to be filled in order to have a productively continuous schema in which it is clear how each step gives rise to the next. During the construction phase of discovery, moving from a sketch to a sufficiently complete schema allows one to work in a piecemeal fashion; one can work on one part of the mechanism at a time

while leaving other parts as black or grey boxes. Because one is attempting to reveal the productive continuity of a mechanism from beginning to end, what one learns about one step of the mechanism places constraints on what likely has come before or what likely comes after a given step.

Abstraction comes in degrees and involves dropping details; specification involves adding details all the way to instantiation, with sufficient details to represent a productively continuous mechanism from beginning to end. A goal in discovery is to find a schema at a given degree of abstraction, from a very abstract type of schema with few specified components to a fully instantiated one for a particular case. For example, the schema DNA→RNA→protein is very abstract. Steps are condensed in this spare representation. However, any given step could be instantiated with specific details if needed for the project at hand. A more detailed schema would begin with a particular coding DNA sequence, show the transcription to complementary messenger RNA, and proceed through the well-known steps of reading the genetic code to order the amino acids in a particular protein.

The desired degree of abstraction depends on the purpose for which the mechanism is sought. Although degree of abstraction is an independent dimension from the scope of the domain to which the schema applies, more abstract schemas (if they have any instances at all) may have a wider scope of applicability. Hence, when the goal of the discovery process is to find a very generally applicable mechanism schema, it is likely to be represented at a high degree of abstraction, as in the above schema for protein synthesis.

The move from how possibly to how plausibly to how actually is driven by applying strategies for evaluation, such as experimental testing, and strategies for anomaly resolution, such as localizing faults and revising the schema. Ideally one wishes to find empirical evidence for each step in the mechanism (Craver and Darden 2013, Chs. 6-9).

Consider the example of mechanisms connecting a gene mutation to a disease phenotype. One starts with the beginning point, e.g., a particular gene mutation, and a characterization of the disease phenotype (e.g., a set of symptoms). At the outset, between the gene/gene mutation and the phenotypic character is a black box. Having evidence of an association between a beginning point (the gene mutation) and the end point (the disease phenotype), the discovery task is to fill in the black box to some degree of detail. For example, if the goal is to replace an identified mutant gene during gene therapy, then it may be unnecessary to find all the intervening steps in the mechanism. A highly abstract schema may be sufficient to guide the work to find and replace the faulty gene. However, if the goal is to design a therapy to alter an entity or activity in a downstream mechanism site, then specific details become important: e.g., one may need to find the three-dimensional structure of a protein and identify its active site or locate the effect of an environmental factor.

A given gene to phenotype mechanism has entities of different size levels, beginning with the macromolecular DNA, proceeding through protein synthesis, which employs ribosomes (particles in the cytoplasm, composed of both proteins and ribosomal RNAs), and on to, in some cases, ever larger level cell organelle, membrane, cell, tissue and organ components. The appropriate size level depends on what the working entities

are in the steps of the mechanism, on how the phenotype is characterized, and how much detail is needed for a given project. Hence, a single gene to phenotype mechanism likely has entities at many different size levels. (For more on the difference between size levels and mechanism levels, see Craver 2007, Ch. 5; Craver and Darden 2013, pp. 21-22)

The mechanism discovery process has at least four aspects: characterizing and recharacterizing the phenomenon, constructing a schema, evaluating the schema, and revising the schema (Darden 2006, Ch. 12; Craver and Darden 2013, Chs. 4-9). These are often pursued in parallel and in interaction with one another. Strategies for mechanism schema construction are the most relevant here. One localizes where the mechanism operates. For gene to phenotype mechanisms, the mechanism starts with a DNA sequence; what the final stage is depends on the characterization of the phenotype. Thus, the overall structure of the mechanism to be discovered begins with DNA and ends with a phenotype. If a library of types of mechanism components is available, then those types of components become candidates to be specialized to construct steps of the target schema. For example, the module of protein synthesis is a likely module to use in an early step in a gene to phenotype mechanism. The strategy of forward/backward chaining allows the mechanism chain builder to reason forward from one step to the following step or backward to a likely previous step. Activity enabling properties in the previous step indicate possible types of mechanism modules to come. Activity signatures indicate what possibly came before, because once a specific kind of activity has operated to change the state of the next step it leaves specific traces (signatures). For example, a polarly charged DNA base is available to bond to its complementary base in the next step of the mechanism of DNA replication. Because hydrogen bonding leaves weakly bonded

molecular components, when such a signature is detected, one can conclude that polar bonding occurred in a previous step (Darden 2002; Darden and Craver 2002; Craver and Darden 2013, Ch. 5).

With this synopsis of some key features of previous work on mechanisms in hand, we now turn to discovery of disease mechanisms. Here too, we argue, the nature of the product guides the reasoning process to find it. The product is a schema representing the steps in a target disease mechanism. The process is the reasoning by a chain builder to construct a diagram to represent the steps, and, while doing so, to fill black boxes to remove uncertainties.

Disease Mechanisms

In medicine, the following general types of mechanisms are of interest:

(i) The "normal" biological mechanism (noting that what is "normal" can nonetheless vary from person to person)

(ii) The general disease mechanism, which aids in finding sites for therapy and designing therapeutic treatments

(iii) The specific disease mechanism in an individual patient, which may aid choosing an effective therapy

(iv) The general mechanism of action of a drug or other therapeutic agent

(v) The specific mechanism of action of drug or therapy in an individual patient, given their genetic makeup and personal history

(vi) Possible mechanisms to account for side effects of therapies on other bodily mechanisms

Philosophers of medicine are participating in a lively debate about the role that knowledge of the mechanisms of the action of therapies (iii-vi above) should play in evidence based medicine. The debated issue is this: is evidence of the effectiveness of a therapy from randomized clinical trials sufficient to show the efficacy of a therapy, or is knowledge of its mechanism of action needed? (See, e.g., Russo and Williamson 2007; Howick 2011; Andersen 2012.) That is not our topic here. Those concerned with evidence for a therapy acknowledge that knowing the disease mechanism(s) (type ii and iii) can aid the rational design of therapies. That is one of our topics here.

The philosopher of science Paul Thagard analyzed reasoning in discovering disease mechanisms and possible therapeutic sites. Diseases are of different types, which he classified according to their causes (Thagard 1999). Thagard noted that one searches for different types of mechanisms if the disease is due to different types of causes. Thagard proposed different types of abstract mechanism schemas, based on the different types of diseases, including infectious disease, nutritional disease, and molecular genetic disease. The causes of diseases, he claimed, are most often identified by statistical and experimental means before researchers find the mechanism in which that cause participates. However, finding the cause and thereby classifying the kind of disease aids the search for the mechanism. In each type, finding where a normal mechanism is broken indicates sites for possible therapeutic intervention (Thagard 1998; 2003).

As to discovering such disease mechanisms, Thagard queried whether Darden's (2002) reasoning strategies for discovering mechanisms might be useful in medicine. This paper shows that they are: schema instantiation, modular subassembly, and

forward/backward chaining are indeed relevant in disease mechanism discovery, as we will see below.

Several philosophers and historians of medicine have discussed cystic fibrosis. Cystic fibrosis (CF) is a monogenic, autosomal (i.e., not sex-linked), recessive (i.e., a patient must have two mutations, one inherited from each parent) disease. The gene (labeled *CFTR*) is large: about 180,000 bases on the long arm of chromosome 7. The CFTR protein transports chloride ions across membranes in epithelial cells. Normal functioning aids in maintaining appropriate salt balance in those cells. Many different types of mutations in this single gene produce variants of the disease. Researchers have identified as many as 1324 different disease causing mutations in the *CFTR* gene (http://www.hgmd.cf.ac.uk/). These mutations produce a cluster of symptoms affecting the lungs, pancreas, liver, and other organs. Lung inflammation and frequent lung infections due to the build up of mucus in the lungs are the most serious problems for CF patients.

The mechanistic analysis applies well to this case. Since the discovery of the relevant gene in 1989, medical researchers have extensively studied the beginning steps in the disease mechanism and have targeted them for therapy (Darden 2013; Craver and Darden 2013, Ch. 11). Sadly, gene therapy has yet to work successfully to put a functioning copy of the large *CFTR* gene into the genome of cystic fibrosis patients (Lindee and Mueller 2011). Downstream stages of the mechanism have proved more promising targets, especially in mechanisms where the mutation produces a malformed protein whose function can be partially corrected with "chaperonin" molecules (Solomon 2015). Despite extensive study, black boxes remain in the later stages of the mechanism.

Still unknown are all the details of exactly how defects in the chloride ion transport mechanism produces the final phenotypic symptoms of the disease in the lungs (Darden 2013). Also puzzling is why patients with the same genetic mutations nonetheless vary in the severity of their symptoms (Solomon 2015).

Cancer is another disease whose mechanisms have been discussed by philosophers. Thagard (2003) classifies cancer as a "disease of cells" due to genetic mutations. Hereditary and somatic mutations occur in oncogenes (genes that regulate cell division or survival) and tumor suppressor genes (suppress cell division). In contrast, the philosopher Anya Plutynski criticizes the view of cancer as merely a genetic disease (Plutynski, forthcoming). Cancer, she says, is a "complex process, due to many causes," not just to gene, chromosomal, and epigenetic changes but also "causes acting at the level of the cell and above" (Plutynski 2013, p. 466). Genetic mutations are often important difference makers in cancer etiology, but they are not the only ones. Genetic models, she argues, inappropriately "black box" environmental factors (Plutynski 2013, p. 474). Mechanism sketches for cancer should depict factors other than genes and indicate where such environmental factors should fill in black boxes.

The product and the discovery process for complex trait diseases, e.g., Crohn's disease, are much more complex than for monogenic diseases and more unknown than for the causes of cancer. The causes of this inflammatory bowel disease are hypothesized to include not only many genetic variants, but also interactions with the intestinal microbiome (the microbes in the gut), and tuning of the immune system through past exposures to invading microbes. So far, Crohn's is statistically associated with over 160 loci in the genome (de Lange et al. 2017). The gene products engage in complex

interactions in producing the disease. Hence, Crohn's researchers need to find many mechanisms connecting gene variants to aspects of the disease phenotype. These will include roles of environmental factors, e.g., diet, and interactions with the gut microbiome. Then researchers need to find complex interactions among the products of the many mechanisms involved in order to explain the disease and find sites for therapies. This paper extends the mechanistic analysis beyond monogenic cases and cancer discussed in previous philosophical work to include these complex trait diseases.

Genome-Wide Association Studies (GWAS) provide data statistically associating genetic variants with disease risk (https://www.ebi.ac.uk/gwas/home). The presence of this association implies that either this variant SNP (single nucleotide polymorphism, a single base change in the DNA) or another variant nearby is involved in a disease mechanism. Many variants are just what may be called "proxy SNPs"; these are single base changes in the DNA that are somehow linked to parts of the genome associated with the disease, but do not themselves play roles in a disease mechanism. The question arises for each identified variant: Is there a disease mechanism that begins with the genetic variant and proceeds to the phenotype, characterized as disease risk?

Discovery of genetic disease mechanisms, we propose, is aided by an abstract diagrammatic representation for disease mechanisms. An abstract diagram sketches the overall structure of the product--the disease mechanism--and thereby aids the chain builder in the process--reasoning to its discovery.

Diagrammatic Representations for Genetic Disease Mechanisms

One analysis of understanding is that it involves the ability to manipulate a mental representation (Wilkenfeld 2013). An abstract diagrammatic representation facilitates the

formation of an individual's understanding, guides the person in filling it in, and serves to convert a single person's visual mental representation into a publically accessible form. The proposed abstract general form for mechanism disease diagrams to be discussed below plays this role. It has several advantages as a representation of the product to guide the process of its discovery:

(a) It provides a general framework for integrating and expanding knowledge about disease mechanisms.

(b) It clearly delineates what is known and not known about the mechanism(s) of each disease.

(c) It provides a potential way of finding interactions when multiple mechanisms interact to produce or modulate the severity of a disease.

(d) It allows representation of interacting subsets of mechanisms (found in (c)) in individual patients.

(e) It facilitates identification of sites of potential therapeutic intervention.

Consider these abstract idealized mechanism diagrams of genetic variant to disease phenotype via disease mechanisms:

Figure 1, The entire mechanism between a genetic variant and disease risk is a black box. The question mark queries whether a mechanism actually exists between the two.

Figure 1 is a beginning point after a genetic variant is related to disease risk: does that variant mark the beginning of a mechanism? In contrast, if indeed a target mechanism is found, then an idealized general abstract diagram for it will have components such as in Figure 2.

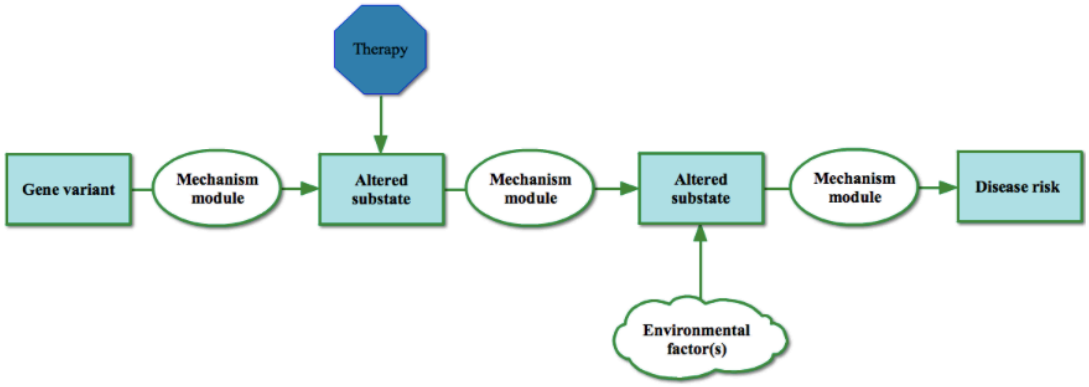

Figure 2, Abstract genetic disease mechanism chain with no black boxes.

Figure 2 shows an idealized diagram of a genetic disease mechanism chain for a case where all the components of the chain are understood. It has no black boxes. The goal of chain building is to proceed from a figure such as Figure 1, to progressively fill the black box, to draw a diagram such as Figure 2 (or else conclude no mechanism likely exists). Figure 2 begins with a variant that affects the function of a gene. Rectangles represent altered substates, with in box text indicating how the substate is altered. Ovals, depicting mechanism modules of groups of activities and entities, label the arrows; text inside the oval names the module. The entities and activities of modules transform one substate to another. The chain proceeds through successive altered substates to a disease phenotype. Blue octagons indicate such potential sites for therapeutic intervention. A white cloud entering the chain from below shows a possible role for environmental factors. All the boxes are glass boxes; one can look inside and see whatever details are relevant. Details in a box or oval may be telescoped (collapsed, as in a folded telescope)

when the details are irrelevant. All the lines are green, indicating the chain builder's highest confidence level, based on evidence for those steps.

Our diagrammatic framework enables us to suggest **a set of heuristic questions**. These serve to guide the chain builder in filling black boxes to remove ignorance and to reach a diagram that is complete enough for whatever is the purpose of the work.

Mechanism at all? The first step in removing ignorance is to inquire whether a mechanism exists at all. In Figure 1, the chain begins with a genetic variant connected via a black box with a question mark to the disease phenotype. For a statistical association, the question mark asks whether there is a mechanism at all. To answer that question: try to fill the black box with a plausible mechanism. Given failure to find a possible mechanism, the chain builder will have to make a judgment call as to when to stop trying. Where a specific mutated gene is known, such as in monogenic diseases and some cancers, then the first box names the gene and its mutation. The chain builder can draw a green arrow to a black box with no question mark to indicate where additional specific mechanism components are expected and should be sought.

What kind of genetic variant begins the chain? Once the task becomes to fill the black box with a mechanism, the next question is what kind of variant begins the chain? Different kinds of variants likely require chains with different kinds of beginning steps. For example, a missense variant (a change in one DNA base that results in a changed amino acid in a protein) will proceed via protein synthesis. In contrast, a variant in a non-

coding region of DNA that affects the binding of a regulatory protein will have earlier steps before the module of protein synthesis plays a role.

In addition to building the chain forward from the genetic variant, is it possible to begin at the end and build the chain backward? Black boxes show missing steps in need of elaboration. Because what comes before and what comes after are indicated in the diagram, the chain builder can reason forward from the previous step or backward from the subsequent one to conjecture what fills the box. Are there activity enabling properties in a step that indicate a likely module and a likely substate perturbation in the next step? Conversely, are there activity signatures (properties of an altered substate) that indicate what kind of activities operated in the previous module, earlier in the chain, that produced it?

Do environmental factors play a role? Is there a place where a white cloud representing an environmental factor should be added? What kind of substate change follows from its insertion?

Does the chain branch into subchains? Does the chain branch at a given step? If so, are the subchains mutually exclusive alternatives ("or" at the branch) or do both occur ("and" at a branch)? Is there uncertainty on the part of the chain builder such that branches should be labeled with "and/or"?

Other than branches in the chain, are there other nonlinear organizational motifs that need to be added? If no feedback or feed-forward loops are included, the question arises as to whether any should be? Are there other nonlinear organizational motifs to consider?

Is there a potential site for therapeutic intervention? Can types of therapies for types of steps be suggested, e.g., does a misfolded protein indicate that a chaperonin should be considered?

How strong is the evidence for each step? How confident is the chain builder in each step? As noted above, a black box with a question mark asks whether there is anything to be discovered at that point, either whether a mechanism as a whole exists or whether a branch of a chain exists. A black box without a question mark indicates a likely but currently unknown substate, mechanism module, or group of substates and mechanism modules. Green, pink and red colors of the lines indicate the confidence level of the chain builder in each specific perturbed substate and arrow/module. Just like black boxes, red and pink colors indicate where more work is needed to increase confidence, to convert red and pink to green. Evidence for particular steps includes the following: standard biological knowledge, one or more reliable published sources provide evidence for the step, and experimental evidence. (The diagrams below illustrate the use of all three kinds of evidence.)

These questions aid the chain builder in using the diagram to depict kinds of ignorance and direct resources to remove it.

Web-based Graphical Interface for Chain Building

We are developing a web-based graphical interface to aid medical researchers in hypothesizing and representing genetic disease mechanisms. The interface is implemented using the conventions discussed above. We have used it to produce more detailed diagrams, such as those below. These three figures provide examples for a monogenic disease chain, a cancer disease chain, and one (of what will be many) complex trait disease chain.

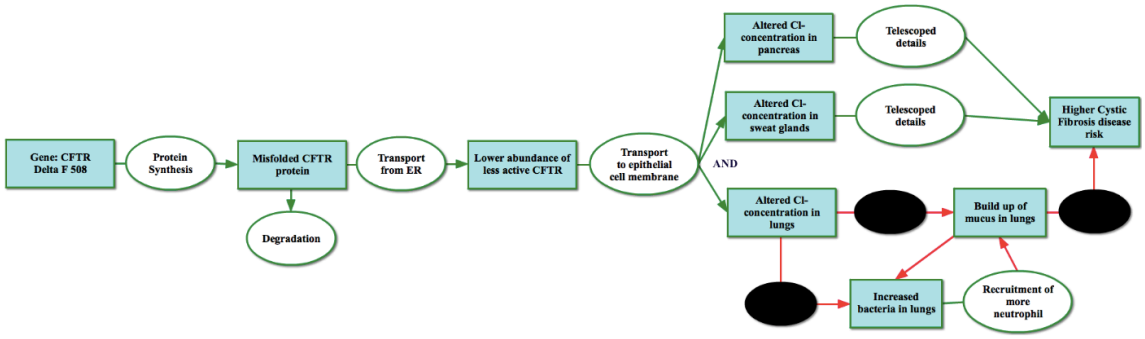

Figure 3, A disease mechanism chain for cystic fibrosis, an example of a monogenic disease.

Figure 3 shows a mechanism disease chain for cystic fibrosis (CF). It begins with the mutation, DeltaF508, in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. This is the most common mutation among CF patients in the United States. Normally the CFTR protein inserts into epithelial cell membranes and transports chloride ions across the membrane. In this mutant form, three DNA bases are missing, resulting in one missing amino acid in the protein. The protein misfolds. The first branch in the chain indicates that some misfolded proteins are degraded but others are released from the endoplasmic reticulum (ER), where it is synthesized. This lower abundance of the protein and its misfolding results in altered concentrations of the misfolded protein in epithelial cells in the pancreas, sweat glands, and lungs, shown in the next three branches. This is a well-studied case so all the lines in the beginning of the mechanism chain are green. The black boxes and red arrows (toward the end in the lower branch of the chain) indicate the controversy that still surrounds exactly what contributes to the build up of thick mucus in the lungs. One hypothesis is that improper salt balance produces the mucus build up

(shown in the top chain coming out of the lungs rectangle). Another hypothesis is that a contributing factor is the break down of the overexpressed immune cells, neutrophils, that are recruited to fight invading bacteria (shown in the loop in the bottom branch of the chain). (For more details, see Darden 2013.)

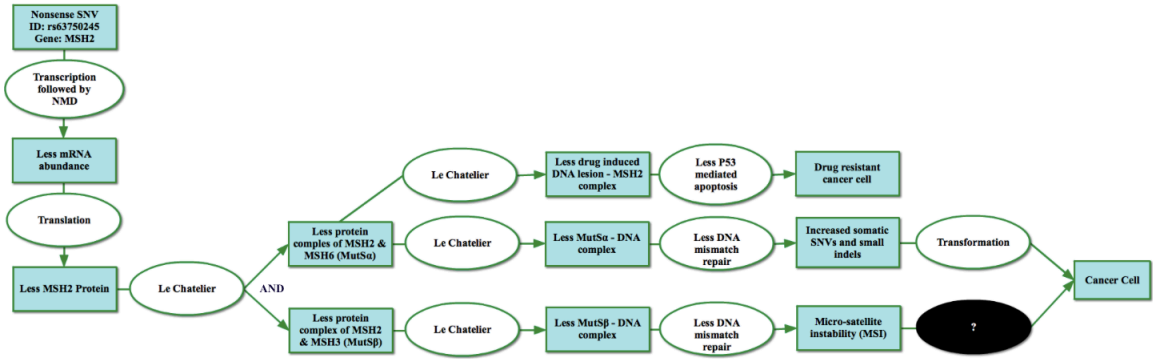

Figure 4, Mechanism chain diagram for a cancer gene variant in the human gene *MSH2*.

Figure 4 is an example of a disease mechanism chain for cancer. This is a hypothesized mechanism chain for a germline DNA variant in the human gene *MSH2*. The ID number identifies the particular variant. The DNA base change is expected to lead to nonsense mediated decay (NMD), decreasing the messenger RNA abundance by half, and as a consequence, also decreasing MSH2 protein abundance. As a result, all the complexes of this protein with other proteins will also be of reduced abundance, hence the "and" at the first branch. The Le Chatelier's Principle refers to a state in which, e.g., the concentration of a reactant changes in a system in equilibrium such that the equilibrium will shift so as to tend to counteract the effect. This activity lowers the abundance of macromolecular complexes in the next steps in all three branches. Then, the branches show the effects of less DNA mismatch repair of both short and longer mismatch regions in the DNA, as well as less apoptosis (programed cell death) triggered by recognition of drug induced DNA damage. Results include greater accumulation of somatic mutations, hence increased cancer risk. Greater microsatellite instability occurs, which may also somehow increase cancer risk, indicated by the black box with a question

mark. The top branch of the chain shows the path to drug resistance. (For details, see the review, Li 2008.)

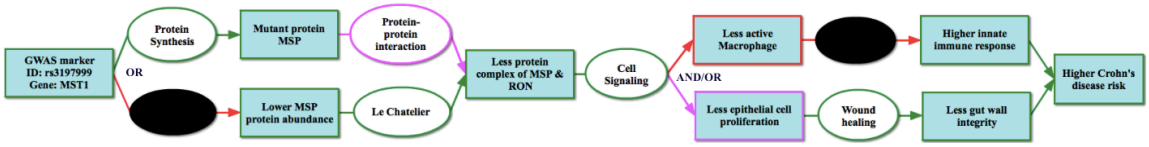

Figure 5 shows a chain for one variant in the *MST1* gene associated with increased risk of Crohn's disease

---

Figure 5 is an example of one of the hypothesized mechanism chains for Crohn's disease, originating in a locus containing a GWAS marker at the *MST1* gene, which codes for MSP (Macrophage Stimulating Protein). The mechanism begins at the perturbed DNA substate on the left, and progresses through protein, protein-protein complex, cell signaling, innate immune response, and gut barrier layer stages to disease risk. In this view, some parts of the chain, at the DNA, protein, and protein complex stages, are fully expanded, while others are partly telescoped (for example "cell signaling" and "innate immunity"). These telescoped steps have multiple substate perturbations and mechanism modules within them. (For details on research on this chain, see Gorlatova et al. 2011.) Black boxes, as well as pink and red lines, indicate uncertainty. The "or" at the first branch indicates two different ways that the chain might branch. The next "and/or" indicates that one or both of the branches may occur; the chain builder is not yet certain which is the case. This *MST1* chain represents just one of the many mechanisms involved

in Crohn's disease. Much work remains to find additional chains and the ways they interact with each other.

This diagrammatic method clearly illustrates the way an abstract representation of the product to be discovered guides reasoning to its discovery through various stages. Admittedly, this diagrammatic representation abstracts away from many features of mechanisms discussed above. It is an open question whether any features of more fully represented mechanisms will need to be added. Note that the diagrams do not include structures of the proteins and protein complexes, although, if needed, it would be easy to add a link to the protein structure database. Also omitted are quantitative rates by which activities operate or quantitative measures of abundance of entities; this is a qualitative representation. Furthermore, the locations of the mechanism steps are not graphically shown, e.g., whether the steps occur in the nucleus, in the cytoplasm, within cell organelles, or elsewhere; however, when relevant, text in the altered substate box does indicate location, such as epithelial cells. It is an open question whether such general features of mechanisms (structures, rates, spatial locations) will need to be represented to fulfill the goals of adequately explaining the disease and locating sites for therapy. Should the need arise, the general philosophical analysis of mechanisms provides a storehouse of items that can be added to the simplified graphic in the future.

Our work contrasts with other graphical forms of representations; we discuss two here. One type is a directed acyclic graph (DAG) to represent causal chains. Philosophers of science are engaged in a lively debate about the adequacy or inadequacy of DAGs for representing normal biological mechanisms (e.g., Gebharter and Kaiser 2014; Kaiser 2016; Weber 2016). From our perspective, causal graphs are impoverished in merely

having unlabeled edges that represent generic cause relations. In contrast, our mechanism diagrams indicate the specific kind of activity or the group of entities and activities in a mechanism module that effect each particular instance of casual production.

Biologists have developed other graphical frameworks, but not (so far as we know) ones using analyses from the new mechanistic philosophy of science to specifically represent genetic disease mechanisms. Most represent normal molecular biological pathways. (For a list see, e.g., Jin et al. 2014.) One of the best developed the Kyoto Encyclopedia of Genes and Genomes. The KEGG Pathway database is a collection of manually drawn graphical diagrams. These represent molecular pathways for metabolism, genetic information processing, environmental information processing, other cellular processes, some human diseases, and drug resistance (Kanehisa 2017). The disease diagrams are represented by perturbations in normal pathways.

The KEGG disease diagrams differ from our framework in numerous ways. Unlike our diagrams, KEGG depicts diseases in pathway wiring diagrams of groups of normal pathways with genes associated with a disease in color-coded rectangles. Furthermore, the focus is only on the early stages involving genes, proteins, and molecular interactions. In contrast, each of our diagrams begins with a specific single gene mutation and traces the changes resulting from that mutation through numerous other stages to the disease phenotype.

For example, in KEGG the Crohn's disease pathway is part of the pathway for inflammatory bowel diseases (IBD) in general. The IBD pathway depicts numerous genes and proteins (in rectangular boxes) in their normal pathways. A few genes known to be associated with diseases (not just Crohn's) are colored pink in contrast to normals, which

are green. Light blue boxes indicate actual drug targets. Some anatomical details are depicted, such as a breach in the gut wall and a macrophage (an immune cell). Types of lines connecting genes and gene products indicate types of interactions, such as inhibition, activation, indirect effect, or dissociation. In contrast, our diagrams have ovals with text to label the arrows, thus showing the activities or mechanism modules that produce perturbed substates. Any new kind of activity easily fits within our framework whereas KEGG will need a new graphical symbol. Consequently, our framework is more easily extendable.

Also unlike ours, the KEGG pathway diagrams do not indicate confidence levels nor do they include black boxes to show ignorance. Our diagrams are thus better for directing the discovery process to produce the product of a genetic disease mechanism by filling black boxes, resolving uncertainties about branches, and increasing confidence levels.

Conclusion

This paper argues for the thesis that the product shapes the process: knowing what is to be discovered provides guidance as to how to discover it. Here the product is a schema to represent steps in a disease mechanism from gene variant to disease phenotype. Heuristic questions and abstract diagrams aid the reasoning process to discover a specific disease mechanism chain. By indicating black boxes and uncertainties, the chain builders represent their ignorance at a given time and show where to focus additional work. This new diagrammatic representational tool, grounded in philosophical analysis, aids in storing collective knowledge and guiding collective discovery.

Plans for future work include finding standardized ontology terms (Arp et al. 2015) for each stage of genetic disease mechanisms (e.g., Gene Ontology 2015). Such standardized terminology is especially important to facilitate finding interactions among the related chains. This standardization will also foster communication between groups of experts to complete the parts of the chains in their areas of expertise. An even longer-range goal is to apply this work in precision medicine. That goal requires finding specific interacting chains for individual patients (or groups of patients), given variability in their genes, environment, and lifestyle, so that personalized therapy can be designed and administered.

References

Abrahamsen, Adele A. and William Bechtel (2015), "Diagrams as Tools for Scientific Reasoning," *Review of Psychology and Philosophy* 6: 117-131.

Abrahamsen, Adele, Benjamin Sheredos, and William Bechtel (forthcoming), "Explaining Visually: Mechanism Diagrams," in Stuart Glennan and Phyllis Illari (eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*.

Andersen, Holly (2012), "Mechanisms: What are They Evidence for in Evidence-based Medicine?" *Journal of Evaluation in Clinical Practice* 18 (5): 992-999.

Arp, Robert, Barry Smith, and Andrew D. Spear, (2015), *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: MIT Press.

Bechtel, William and Adele Abrahamsen (2005), "Explanation: A Mechanist Alternative," in Carl F. Craver and Lindley Darden (eds.), Special Issue: "Mechanisms in Biology," *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 421-441.

Bechtel, William and Adele A. Abrahamsen (2013), "Thinking Dynamically About Biological Mechanisms: Networks of Coupled Oscillators," *Foundations of Science* 18: 707-723.

Bechtel, William and Robert C. Richardson (1993), *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

Craver, Carl F. (2007), *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York: Oxford University Press.

Craver, Carl F. (2008), "Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential," *Philosophy of Science* 75 (5): 1022-1033.

Craver, Carl F. and Lindley Darden (2013), *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago, IL: University of Chicago Press.

Darden, Lindley (2002), "Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining," *Philosophy of Science* 69 (Proceedings): S354-S365.

Darden, Lindley (2006), *Reasoning in Biological Discoveries: Mechanisms, Interfield Relations, and Anomaly Resolution*. New York: Cambridge University Press.

Darden, Lindley (2013), "Mechanisms versus Causes in Biology and Medicine," in Hsiang-Ke Chao, Szu-Ting Chen, and Roberta L. Millstein (eds.), *Mechanism and Causality in Biology and Economics*. The Netherlands: Springer, pp. 19-34.

Darden, Lindley and Carl F. Craver (2002), "Strategies in the Interfield Discovery of the Mechanism of Protein Synthesis," *Studies in History and Philosophy of Biological and Biomedical Sciences* 33: 1-28. Reprinted with corrections in Darden (2006, Ch. 3).

de Lange, Katrina M. et al. (2017), "Genome-wide Association Study Implicates Immune Activation of Multiple Integrin Genes in Inflammatory Bowel Disease," *Nature Genetics* 49: 256–261. doi:10.1038/ng.3760

Gebharter, Alexander and Marie I. Kaiser (2014), "Causal Graphs and Biological Mechanisms," in Marie I. Kaiser, Oliver R. Scholz, Daniel Plenge, Andreas Hüttemann (eds.), *Explanation in the Special Sciences*. Dordrecht*:* Synthese Library Volume 367, pp. 55-85.

Gene Ontology Consortium (2015), "The Gene Ontology Consortium: Going Forward," *Nucleic Acids Research* 43 (database issue): D1049-D1056.

Glennan, Stuart S. (1996), "Mechanisms and The Nature of Causation," *Erkenntnis* 44: 49-71.

Glennan, Stuart and Phyllis Illari (eds.) (forthcoming), *Routledge Handbook of Mechanisms and Mechanical Philosophy*.

Gorlatova, Natalia, Kinlin Chao, Lipika R. Pal, Rawan Hanna Araj, Andrey Galkin, Illarion Turko, John Moult, Osnat Herzberg (2011), "Protein Characterization of a Candidate Mechanism SNP for Crohn's Disease: The Macrophage Stimulating Protein R689C Substitution," PLOS ONE (open access) http://dx.doi.org/10.1371/journal.pone.0027269

Howick, Jeremy (2011), "Exposing the Vanities—and a Qualified Defense—of Mechanistic Reasoning in Health Care Decision Making," *Philosophy of Science* 78: 926-940.

Jin, Lv, Xiao-Yu Zuo, Wei-Yang Su, Xiao-Lei Zhao, Man-Qiong Yuan, Li-Zhen Han, Xiang Zhao, Ye-Da Chen, and Shao-Qi Rao (2014), "Pathway-based Analysis Tools for Complex Diseases: A Review," *Genomics, Proteomics & Bioinformatics* 12 (5): 210-220. doi:10.1016/j.gpb.2014.10.002.

Kanehisa, Minoru, Miho Furumichi, Mao Tanabe, Yoko Sata, and Kanae Norishima (2017), "KEGG: New perspectives on Genomes, Pathways, Diseases and Drugs, " *Nucleic Acids Research* 45 (D1): D353-D361. https://doi.org/10.1093/nar/gkw1092

Kaiser, Marie I. (2016), "On the Limits of Causal Modeling: Spatially-Structured Complex Biological Phenomena," *Philosophy of Science* 83 (5): 921-933.

Li, Guo-Min (2008), "Mechanisms and Functions of DNA Mismatch Repair," *Cell Research* 18:85–98.

Lindee, Susan and Rebecca Mueller (2011), "Is Cystic Fibrosis Genetic Medicine's Canary?" *Perspectives in Biology and Medicine* 54 (3): 316-331.

Machamer, Peter, Lindley Darden, and Carl F. Craver (2000), "Thinking About Mechanisms," *Philosophy of Science* 67: 1-25.

Plutynski, Anya (2013), "Cancer and the Goals of Integration," *Studies in the History and Philosophy of Biological and Biomedical Sciences* 4: 466-476.

Plutynski, Anya (forthcoming), *Explaining Cancer: Philosophical Issues in Cancer Causation, Evidence and Explanation*.

Russo, Federica and Jon Williamson (2007), "Interpreting Causality in the Health Sciences," *International Studies in the Philosophy of Science* 21 (2): 157-170.

Sheredos, Benjamin, Daniel Burston, Adele Abrahamsen, and William Bechtel (2013), "Why Do Biologists Use So Many Diagrams," *Philosophy of Science* 80 (5): 931-944.

Solomon, Mariam (2015), *Making Medical Knowledge*. New York: Oxford University Press.

Thagard, Paul (1998), "Explaining Disease: Causes, Correlations, and Mechanisms," *Minds and Machines* 8: 61-78.

Thagard, Paul (1999), *How Scientists Explain Disease*. Princeton, NJ: Princeton University Press.

Thagard, Paul (2003), "Pathways to Biomedical Discovery," *Philosophy of Science* 70: 235-254.

Weber, Marcel (2016), "On the Incompatibility of Dynamical Biological Mechanisms and Causal Graphs," *Philosophy of Science* 83 (5): 959-971.

Wilkenfeld, Daniel (2013), "Understanding as Representation Manipulability," *Synthese* (2013) 190: 997–1016. DOI 10.1007/s11229-011-0055-x

---

Acknowledgments. LD thanks Emiliano Ippoliti and his colleagues at Sapienza University in Rome for the invitation to the Building Theories Workshop, and to them and all the participants for lively discussions of discovery heuristics. For helpful suggestions on earlier drafts, she thanks Carl Craver, Giamila Fantuzzi, Nancy Hall, and an anonymous reviewer, as well as Kal Kalewold and the other students in her graduate seminar on mechanisms and ontology. This work was supported in part by NIH (US National Institutes of Health) grants R01GM102810 and R01GM104436 to JM.

Affiliations of authors:
Lindley Darden, Department of Philosophy, University of Maryland College Park, mailing address: 1102A Skinner, 4300 Chapel Lane, College Park MD 20742 USA; corresponding author: darden@umd.edu

Lipika R. Pal, Institute of Bioscience and Biotechnology Research, University of Maryland College Park

Kunal Kundu, Institute of Bioscience and Biotechnology Research; and Computational Biology, Bioinformatics and Genomics, Biological Sciences Graduate Program, University of Maryland College Park

John Moult, Institute of Bioscience and Biotechnology Research; and Department of Cell Biology and Molecular Genetics, University of Maryland College Park