Studies in History and Philosophy of Biological and Biomedical Sciences 44 (2013) 466–476

Contents lists available at SciVerse ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

# Cancer and the goals of integration

Anya Plutynski

*University of Utah, Department of Philosophy, 215 S. Central Campus Dr., 402 CTIHB, Salt Lake City, UT 84112, United States*

ARTICLE INFO

*Article history:*
Available online 10 April 2013

*Keywords:*
Cancer
Integration
Unification
Reduction
Causation
Modeling

ABSTRACT

Cancer is not one, but many diseases, and each is a product of a variety of causes acting (and interacting) at distinct temporal and spatial scales, or "levels" in the biological hierarchy. In part because of this diversity of cancer types and causes, there has been a diversity of models, hypotheses, and explanations of carcinogenesis. However, there is one model of carcinogenesis that seems to have survived the diversification of cancer types: the multi-stage model of carcinogenesis. This paper examines the history of the multistage theory, and uses the theory as a case study in the limits and goals of unification as a theoretical virtue, comparing and contrasting it with "integrative" research.





## 1. Introduction

In this paper, we will consider one central project in the history of cancer[1] research, modeling carcinogenesis as a multi-stage process, as a case study for investigating the ideals of "unification" versus "integration" in the sciences. Multistage models of cancer represent cancer initiation and progression to neoplastic state as a multi-stage process, driven by the acquisition of a series of mutations. Sometimes this view is assimilated with an "evolutionary" perspective on cancer, since a cancer's capacity to attract a blood supply, invade neighboring tissue, and metastasize, are all seen as the result of the acquisition of a series of mutations that increase the relative "fitness" of the cancer cells (Merlo, Pepper, Reid, & Maley 2006; Nowell, 1976). Whether or no cancer progression is best viewed as an evolutionary process is a question requiring further exploration (Plutynski, in press); so, we focus here on the multistage theory.

Cancer incidence increases as a power of age; the multistage theory explains this phenomenon as due to the rate-limited accumulation of mutations to genes (as well as chromosomal and epigenetic changes) that play key roles in the regulation of the cell cycle. The theory also explains departures from average age of incidence curves. For instance, familial forms of cancer cause a shift the age of incidence curves, due to hereditary mutations that "accelerate" the onset of cancer (Knudson, 1971). The multistage theory appears to explain some patterns of cancer incidence quite well. Most notably, for colon cancer, not only have the specific series of mutations leading to a specific cancer type have been identified, but their mechanisms of action, and thus role in causing dysplastic growth, are well understood (Fearon & Vogelstein, 1990). The history of the multistage theory is a useful illustration of both the advantages and limits of mathematical modeling in arriving at general theories in biomedicine.

Carcinogenesis is a complex process, due to many causes acting both at the level of the cell and above (Bissell & Short, 2009). How, if at all, may a simple mathematical model capture all the various causes of cancer(s), acting at distinct temporal and spatial scales? It cannot, and it should not. Abstract models, such as the family of models of carcinogenesis, are intended to identify the central causal factors yielding some outcome, at one well-defined level

*E-mail address:* plutynski@philosophy.utah.edu

[1] One reviewer asks that I "define cancer" in a sentence or two. In my view, this is very difficult to do, first, because cancer is many different diseases (see, e.g., NCI's website on "Defining Cancer," http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer). Second, like many biological kinds, cancer is, at best, a HPC kind (Williams, 2011); there are properties that many cancers share, but no necessary and sufficient conditions or essential properties. At minimum, cancer is a breakdown in features of the organism that control cell birth and death; these include not only mutations, but also chromosomal changes, epigenetic factors, tissue disorganization, and much else besides. Addressing the variety of potential definitions of cancer and their limits would require a very different paper. I address this and related issues in a forthcoming book.

1369-8486/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.shpsc.2013.03.019

of analysis. So, they deliberately exclude, for instance, causes exogenous to the system of interest. Such causes are treated as more or less a black box. At least initially, there was no representation in the multi-stage models of the role of the tumor microenvironment, immune system, diet, or smoking. Nonetheless, the models were a way of systematically representing carcinogenesis, consistent with a variety of independent evidence: patterns of cancer incidence by age, patterns of cancer incidence in childhood cancers, and toxicological data on the effects of chemical carcinogens on animals. More recently, such models have been integrated with data from molecular genetics on the role of specific genes in cell birth and death, and data on the rate and structure of cell renewal in different tissue types (Frank, 2007; Frank & Nowak, 2004). That is, what began as a way of modeling cancer at one level of analysis using simple mathematical models became a theoretical framework for integrating new data from different levels of analysis—both from the "bottom up" and "top down".

Biologist Steven Frank (2007) calls the family of models that represent cancer as a multi-stage process the "dynamics" of cancer. This characterization suggests an analogy with Newtonian dynamics, the theory that unified terrestrial and celestial mechanics. Mathematical representations of cancer initiation and progression as a dynamic, multi-stage process are analogous to Newtonian mechanics in the following respects. They both treat complex phenomena using simple mathematical models, and both hypothesized that there were common causes, driving observed patterns. Moreover, both treat different phenomena—in the case of multistage theory, distinct cancers—as of a kind. Hereditary and somatic cancers, cancers found in different tissues or of different types are all, on this theory, subject to similar causes acting in similar ways. This unifying perspective had the virtue that it served as a guiding idea for a research program. Seeing distinct cancers as subject to similar causes was central to a research program that (in part) led to the discovery of a family of genes that play important roles in all cancers: *TP53*, *RB*, *APC*, *HRAS*. It also led to the realization that understanding genes and their activity in isolation from the tumor microenvironment was not sufficient to explain carcinogenesis. Thus, the history of multistage theory can serve as an interesting case study for the purported virtues of "integrative" or "unified" theories and explanations in the sciences.

As many philosophers writing on the explanatory power of unifying theories have noted (see Cartwright, 1980) unification often comes at a cost; unified theories or laws with wide scope trade generality and cohesiveness for simplification or omission of complex causal details. And, if anything is an instance of a complex causal processes, carcinogenesis is it. No single model could possibly incorporate all the factors affecting carcinogenesis; in part, because cancers are so different, but in part, also, because cancer is not simply a "genetic" disease. Mutations are a significant difference maker in cancer, but they are not the only one (Bissell & Hines, 2011). The multistage theory focuses causal factors that shift the age of incidence curve: core difference makers to the time of onset of cancer. This accumulation of mutations is taken to explain the fact that cancer incidence by and large increases as a power of age.

By focusing on mutations, the multistage theory trades simplicity and unifying power for explanatory detail. However, over time, the theory has come to incorporate evidence from a wider domain. In other words, the multistage model of cancer is a case study in integration as a process. While in some sense, it started as a "reductive" and "unifying" theory, reducing carcinogenesis to nothing more than the serial acquisition of genetic mutations, over time, it has incorporated more data, from a variety of methodological and theoretical perspectives. Though, even in the beginning, the theory required moving between levels of analysis; that is, it required the insight that epidemiological data on age of incidence might provide some clue as to the etiology of cancer.

### 1.1. Cancer and the call for integration

While the sheer number and diversity of questions and subject matter in cancer research would seem to argue for greater specialization, recently, there has been a call from both granting agencies and major research universities for more "integrative" and "inter-" or "trans-disciplinary" research.[2] As an example, in 2003, the US National Cancer Institute's (NCI) Division of Cancer Biology "initiated a program to highlight (1) systems biology, (2) a systems approach to cancer biology, (3) interdisciplinary and collaborative research and (4) interdisciplinary training." With an initial $14.9 million in funding, the Integrative Cancer Biology Program (ICBP) was created, and in 2004, nine interdisciplinary centers were founded, "incorporating a spectrum of new technologies such as genomics, proteomics, and molecular imaging, to generate computer and mathematical models that could predict the cancer process." (NCI, 2012, http://icbp.nci.nih.gov/). "Integrative" research, in part as a result of funding initiatives such as the above, has been promoted in many other areas in biomedical science: many institutions have founded research programs, institutes, and centers of "integrative" research. A search in PubMed with the terms "cancer" and "integration" turns up over 9000 hits.

What exactly is being called for with these demands for integrative research? And why should we presume that integration would be a good thing? Perhaps because so many kinds of things may be "integrated" (data, methods, explanations; see O'Malley, this issue), there are many different meanings at play; and, all too often the term is used rhetorically as an advertisement of forward-thinking science with very little warrant. So, what *does* warrant the appellation? This paper will provide an overview of different philosophical accounts of both what integration might be (Section 2); a case study of the multistage theory (Section 3); and, finally an argument for the following: (a) the success of explanatory "integration" is always relative to some specific scientific problem (cf. Brigandt, 2010, 2013; also, Love, 2008), and so one should be wary of generalizations about "*the*" goals of explanatory integration, and (b) theoretical frameworks may become successively more "integrative" over time, along a variety of dimensions (Section 4). While these points may be not entirely novel (see, e.g., Bechtel, 2010; Brigandt, 2013; Love, 2010), the below will hopefully provide some framework for future discussion of both the goals of integration in future philosophical work, particularly in the context of biomedical research.

## 2. What is integration? How is it distinct from unification?

A variety of philosophers of science have offered models of "integrative" research, and "integrative explanation" in

[2] The NSF, the National Academy of Sciences, the MacArthur Foundation, the Mellon Foundation, and the Robert Wood Johnson Foundation, have all issued calls for more interdisciplinary, "integrative" research (Brint, Turk-Bicakci, Proctor, & Murphy, 2009; Kessel et al., 2008; National Academy of Sciences, 2005). Most relevant to cancer, however, the NCI founded the "Integrative Cancer Biology Program: Centers for Cancer Systems Biology," which set up several institutions as centers for study of cancer as a "complex biological system," including Columbia, MIT, Georgetown, Memorial Sloan-Kettering, and UT Austin. Most of these institutions focus on computational modeling of cancer dynamics and genomics. In 2009, 2010, and 2011, the NIH issued a series of funding opportunity announcements supporting "new collaborative projects between investigators associated with the Integrative Cancer Biology Program (ICBP) or Tumor Microenvironment Network (TMEN) and researchers who are not involved with the program with which they propose to collaborate" (http://grants.nih.gov/grants/guide/pa-files/par-09-026.html).

overlapping problem domains, as well as "interfield" theories, as one alternative to the view that unification in the sciences is achieved viz. theory reduction in Nagel's (1961) sense (cf. Bechtel, 1986, 1993, 2010; Brigandt, 2010; Brigandt & Love, 2012; Craver, 2005, 2007; Darden, 2005; Darden & Maull, 1977; Love, 2008, 2010; Maull, 1977; Mitchell, 2003; Mitchell & Dietrich, 2006; O'Malley & Soyer, 2011). What exactly is being "integrated," how, and why is this (presumed to be) a good thing? Further, how is integrative research distinct from "unification" as a goal in the sciences? In this section, I will first introduce an *analytic framework* for discussing integration (explicating, in turn, research traditions, problem domains, disciplines, standards, methods, and objects of explanation as these terms are used in the literature in integration), provide a brief overview of a few of the distinctive features of integrative research that the above authors have identified, review some of the motivations for integration, and then turn to distinctions between integration and unification.

First, a research tradition, following Laudan (1977) is "a set of general assumptions about the entities and processes in a domain of study, and about the appropriate methods to be used for investigating the problems and constructing theories in that domain" (p. 81). Such a research tradition is successful when it leads to "solutions" to "an increasing range of empirical and conceptual problems." Success is not necessarily due to the truth of its component theories or accuracy in its ontology, however; Laudan argues that a research tradition can be "fruitful in generating theories" yet "flawed in its ontology or methodology" (cf. p. 82). Moreover, research traditions have a historical dimension, and can change over time; so, it is mistaken to view them as "true" or "false," in the way that hypotheses may be. Finally, and more to our purposes here, Laudan speaks of research traditions being or becoming "integrated," and such amalgamation of traditions may "suggest important new lines of research, and put scientists in a position to deal with empirical and conceptual problems which neither of the ancestor traditions alone could resolve satisfactorily" (p. 104). Laudan gives several examples of research traditions in the physical sciences, (Aristotelian, Galilean, Cartesian, Newtonian), and describes changes in them over time. As historians of science have pointed out (McMullin, 1979), such traditions are not discrete—they both merge and change over time (even embracing contradictory assumptions, a point the Laudan grants), so it becomes difficult to demarcate them, or say when one has been replaced by another.

Contemporary philosophers of science working on integration, perhaps sensitive to this concern, have adopted variants on Laudan's views. For instance, one of the first models of "integration," Darden and Maull's notion of an "interfield theory," treated the units of integration as "two previously unrelated fields," where fields are defined by a set of common questions, techniques, and methods. In one of Darden and Maull's examples, the fields in question were Mendelian genetics and cytology, which were integrated via the chromosomal theory of inheritance. Today, it is common to find philosophers of science speaking of "problem domains" or "problem agendas" as loci of integration (Brigandt, 2010, 2013; Love, 2008), where what defines these domains is a set of questions, and related criteria of explanatory adequacy. For instance, one of Love's central examples is the inquiry into the origin of evolutionary novelties. Explaining the origin of novel morphological structures in a species, qualitatively different from ancestral traits, is a problem agenda that may be broken down into a variety of questions, and answering such questions requires drawing upon a variety of concepts and methods from several disciplines.

Drawing upon this analytic framework, then, the features of integration that a variety of authors mentioned above have identified are the following:

- Integration involves some form of collaboration between fields or research programs, ranging from drawing upon evidence from different fields (a minimum requirement), to sharing laboratory techniques, methods, or concepts, on up to placing constraints on explanation from one field to another.
- Researchers studying the same subject or question, but from different traditions or with different methodologies, might be called upon to collaborate or to make connections between their different approaches.
- Integration frequently occurs at historical periods when scientists seek to address new questions or problems that require information or tools that might fall outside of their discipline, or even any existing discipline (see, e.g., Bechtel, 1986, 1993, Brigandt, 2010; Darden & Maull, 1977; Love, 2010).
- Integration may occur both "within" and "across" "levels" of analysis, or distinct spatial, or temporal scales (Craver, 2005, 2007; see also, Mitchell & Dietrich, 2006).
- Integrative explanations are frequently local; that is, they often involve establishment of epistemic and ontological connections between two specific areas of research, (e.g., cancer epidemiology and molecular genetics of cancer), rather than, e.g., two entire disciplines (e.g., physics and chemistry).
- Integration is often contrasted with "unification," in that it is not concerned to provide a single theoretical framework for a whole field or discipline, but rather solve a specific problem, or address a particular question (see, e.g., Mitchell, 2003).
- The nature of integration is such that it may involve "transient" connections between otherwise independent domains of research, targeted at addressing specific questions or problems (Brigandt, 2010).
- A few philosophers working on integration take the search for "mechanism" as a central factor driving integrative work (see, e.g., Bechtel, 1986, 1993, 2010, this issue; Craver, 2007; Darden, 2005).
- Those emphasizing a search for mechanism often associate integration with the placing of "constraints" on explanations at one scale of analysis by another. Such constraints may be "top-down"—e.g., a description of system in which a mechanism resides will determine what component parts or mechanisms will be relevant in some explanation—or, "bottom up" in the sense that discoveries about lower-level phenomena may constrain the "higher level" explanatory possibilities as well (see, e.g., Bechtel, 2010; Craver, 2007).
- All advocates of integration see it as an alternative to the program of theory reduction via deductive subsumption, as described by Nagel (1961), though some defend "explanatory" and "mechanistic" reduction as not only consistent with but an important component of integration (see, e.g., Bechtel & Hamilton 2007).
- Work on integration has tended to emphasize the institutional and social factors involved in making connections between disciplines. Thus, e.g., Darden and Maull's account of "interfield" theories emphasized expectations, techniques and methods over "concepts, laws or theories," (Darden & Maull, 1977) and Bechtel (1993) emphasizes the role of laboratory cultures, and how the institutional features of science function either to promote or fail to promote integrative work (see also Gerson, this issue).

It seems that what all those defining integration share is the notion that what matters is not only *that* researchers share data, methods, concepts, laboratories, tools etc., but *how* this process of collaboration shapes their research. At minimum, the collaborative activity must lead either to (a) a transformation in researchers' conception of the object of explanation, or (b) a change in what would count as an adequate explanation of that object, or (c) both. One way of cashing this out is in terms of placing "constraints" on

accounts of the object, or explanations, from one field or domain or research tradition by another. For instance, those who share a commitment to the view that discovery of mechanisms is central to scientific explanation will (not surprisingly) see explanatory integration as facilitated by the discovery of mechanisms, or, as Craver argues in the context of neuroscience, "constraints on mechanistic organization . . . act as loci for interfield integration" (Craver, 2007). For instance, information about different levels or component parts of a mechanism may constrain explanations at other levels.

Why is integration a good thing? Is it always to be sought? How is it distinct from the virtue of "unification"? Are the two goals necessarily at odds? The current thinking seems to be that the virtue of integration is distinct from that of "unification" in that it is problem-oriented. Bechtel (1986, 2006) argues that integration (e.g., of disciplines, methods, or laboratory cultures) is primarily in service of solving a specific problem. Different disciplines (e.g., chemistry, genetics) or perhaps subdisciplines (akin to "fields," e.g., molecular genetics of cancer), will deploy their different tools and conceptual frameworks together to resolve a problem that may not be resolvable by any discipline in isolation. Thus, integration is a virtue of a kind of "meta-methodology," in service of solving a specific kind of problem (see O'Malley, this issue). Integrationists may be pluralists about explanation; the world (and our scientific explanations of it) may never be wholly integrated; there may be very different (and perhaps equally serviceable) ways of breaking up world that may (in the end) be incompatible.[3] Nonetheless, integration may be both possible and desirable for specific problems domains where there is meaningful overlap between the entities and properties of concern to different domains or subdisciplines. Integrative work, on this picture, might not lead to greater unity of the sciences as a whole, and may be compatible with a modest pluralism (see, e.g., Kellert et al., 2006; Mitchell, 2003).

In contrast, "unification" is usually understood to be a regulative ideal for "theories" or "explanations" (Grantham, 2004; Kitcher, 1999). More "unified" theories are taken to be more comprehensive, or "cover" a wider scope of phenomena. A more unified or unifying theory is a theory that thus permits inferences about and explanations of diverse kinds of phenomena, drawing upon common causes, processes, or "explanatory schema" (Kitcher, 1981). According to Strevens (2008), the unificationist account of explanation is an instance of the "pattern-subsumption" account; where, a phenomenon is explained via subsumption under a general model, explanation schema, or law. Such an account is not necessarily at odds with a causal account of explanation. For instance, Strevens (2004, 2008) argues for a "karietic" account of explanation, according to which a theory that "unifies" is to be preferred when and if it picks out the "shared difference-making" factors, or the most salient causes of distinct phenomena. "Unification" is standardly attributed to theories, models, or explanations, whereas integration is usually attributed to a problem-solving activity or research program that requires sharing data, methods or theoretical resources.

However, integration and unification do have some features in common. For instance, both often result in a transformation in the object of explanation. In the case of integration, distinct fields may shed new light on a problem or object of study. In the case of unification, objects of explanation previously seen as distinct or unrelated may be understood to be of a kind, subject to similar causes, or common explanatory schemes. Second, integration and unification both involve transformation of what would count as an adequate explanation of said object; both unifying theories and integration reconfigure explanations across distinct domains, though through different means. Third, while unification is usually seen as an occurring once, whereas integration is an iterated process, sometimes taking decades of exchange between neighboring disciplines, we will argue that a unified theory or model may be the first step in a more integrative research program.

The two goals may seem to pull in different directions, as unification is often associated with reduction. On one classical view of unifying theories (Friedman, 1974), we unify when we "reduce" the number of phenomena we must take as "brute". However, in our view, sometimes a unified theory may become the framework for a more integrative research program, as we will argue in the case of the multi-stage model of cancer. There are of course many other dimensions along which one might view scientific activity as more or less "integrative" (e.g., conceptual integration, institutional integration, etc.), but for the purposes of the case study to follow, we will follow O'Malley and Soyer's (2011) focus on data integration, methodological integration, and explanatory integration.

Let us take a minimal condition on data integration to be sharing evidence. More stringent conditions on data integration would involve synthesizing or making comparable data types from very different sources, or discerning patterns among data (Leonelli, this issue). As O'Malley and Soyer (2011) describe this process in contemporary molecular biology, this often involves "theorizing and modelling databases, quantifying data accurately, developing standardization procedures, cleaning data, and providing efficient and user-friendly interfaces to enable data not only to be reused, but reanalyzed and combined in novel ways" (p. 61). Examples of this kind of data integration are to be found today in climate modeling, bioinformatics, and molecular genetics and genomics, where a vast influx of data needs to be compared or shared. The creation of shared databases, with standardized ontologies, is no simple task, but one which often transforms or reconfigures the object of explanation in a way that can yield novel hypotheses or explanations (O'Malley and Soyer, 2011).

A second dimension along which integration can occur is by sharing or integrating methods or tools of analysis. There are both *specific* methods or tools for addressing specific questions or problems (e.g., Luria-Delbruck fluctuation test, the McDonald-Kreitman test), and very general methodologies (e.g., the iterated process of mathematical modeling of a dynamic). A minimum condition on methodological integration is that the tools are imported or exported from one subdiscipline or domain to another; a more stringent condition is that the process of sharing and developing such tools informs both what and how the object of investigation is conceived (e.g., as a dynamic process), and what might constitute adequate explanations of the object. One way in which this has been done in the past is when mathematical representations are imported from one into another context; this happened with Maxwell's adoption of a dynamical theory for the behavior of molecules in a gas, suggesting the use of a set of equations for representing their behavior that he might not have otherwise (Achinstein, 1987). The development of formal or theoretical models can be one way of delimiting the "possible" and the actual (see, e.g., Lewontin, 2000), given some set of assumptions about the system of interest. Such development is iterated, or often involves multiple rounds of integrating new evidence, rethinking or qualifying assumptions of the model(s), thus reconceptualizing both the object and the best modeling strategy for that object. Greater methodological integration also might involve the use or synthesis of multiple modalities for modeling the same system, process, or entity. For instance, a simple dynamical model that represents change over time could be used as a first step toward a three dimensional simulation, one which gives a more comprehensive

[3] Some see this as an open empirical question, others take it as a matter of fact; for a discussion, see, e.g., Kellert, Longino, & Waters (2006).

picture of the phenomenon modeled (see, e.g., Spencer, Gerety, Pienty, & Forrest, 2006, for a multistage three-dimensional agent-based model of cancer progression).

Explanatory integration involves answering a novel question, in a way that draws upon multiple different sources of evidence from different fields. This could come about by showing how a pattern investigated in one field is a product of a process or mechanism investigated by a different field, or, simply establishing the consistency and/or complementarity of theory, models, hypotheses, or data of different types. For instance, Fisher (1918) demonstrated the consistency of continuous characters with a particulate theory of inheritance, thus resolving a perceived conflict between Darwinism and Mendelism (Plutynski, 2004). That is, explanatory integration does not require identification of a mechanism, though mechanisms can be useful tools for fostering integration. A mathematical model, computer simulation, or other forms of representation can synthesize data in a new way, showing how patterns or processes at one level are linked to or consistent with patterns or processes at another. This kind of synthesis is explanatory, even if the explanation captures one part or aspect of a dynamics of some system (see also Brigandt, this issue). Successive iterations of models might lead to the transformation, or reshaping, of one model or conceptualization of the system of interest by another. An example is the gradual transformation of the Lotke equations of population growth from a simple logistical representation of rate limited change to contemporary versions with multiple species, non-overlapping generations, etc. (see, e.g. Kingsland, 1995 for a history). This gradual transformation involved explanatory reframing of the forces governing change in populations via incorporation of better data, better modeling strategies, and, over time, a better conceptualization of the object of explanation. This is an example of greater integration leading away from simplicity toward greater complexity (Mitchell, 2003).

There are many examples of each of each of the above kinds of integrative work in cancer biology today—from the construction of databases in cancer genomics, to the development of "systems" epidemiology, the representation of a suite of interacting causes of cancer from the molecular to behavioral (Hoos et al., 2011; Spitz & Bondy, 2010). These projects are integrative in a variety of different senses: they may integrate data, methods, or explanations drawing upon fields of research that have previously been independent. However, the focus of this paper, and the following section, will be a single case study, the multistage model of cancer's dynamics. How, if at all, is this family of models "integrative"?

## 3. Case study: multi-stage model of carcinogenesis

Starting in the 1940s, biologists began to develop a family of models that represent cancer initiation and progression mathematically as a product of a series of rate-limited steps. The first quantitative mathematical models to represent cancer as the product of multistage progression were developed by two biologists who found that mice acquired skin tumors after repeated application of benzopyrene (Charles & Luce-Clausen, 1942). Charles and Luce-Clawson hypothesized that cancer was the product of a "series of mutations," to a single gene that led to the development of dysplasia or atypical growth of cells. Over the next several decades, Armitage and Doll (1954), and Knudson (1971) elaborated upon these mathematical/biological models of cancer as a multistage process. Some of these focused on single genes, some on multiple genes; some traced age-specific incidence to two "hits," others traced it to five or six "hits." All drew upon a variety of different sources of data about cancer: epidemiological data showing patterns of cancer incidence by age, data on age of inheritance in childhood cancer, toxicological data on the effects of chemical carcinogens on animals, and most recently, data from molecular genetics (Balmain, 2001), as well as data on the rate and structure of cell renewal in different tissue types (Frank, 2007).

As with all mathematical models in biology, the multistage models of cancer progression are only a partial representations of the complex process of carcinogenesis, though one that has been enormously fruitful in both framing hypotheses and integrating data from different research programs. One of the major domains of growth in what gets called "integrative" cancer research (or at least one of the major sources of funding by programs like ICBP) is mathematical and computational models to represent the complex causes yielding cancer (Galea, Riddle, & Kaplan, 2010). These models have become enormously sophisticated, representing complex networks of causes and whole systems of interaction and feedback both within and across scales (see, e.g., Spitz & Bondy, 2010). The multistage theory was one of the first attempts to generate a unified mathematical model of cancer initiation and progression to neoplastic state. While the early models were simple, idealized representations, they have become more complex, integrating a wider array of data, over time. In this section, we will apply two of the three dimensions mentioned above to the case of models of cancer dynamics. There are two ways in which modeling carcinogenesis is currently becoming more "integrative":

- First, cancer dynamics integrates *data* from different subdisciplines: cell and molecular biology, developmental biology, epidemiology, and medicine (particularly research into patterns of incidence familial or "hereditary" cancer), and molecular genetics.
- Second, cancer dynamics constitutes an *explanatory* integration of patterns of cancer incidence and dynamics of cancer initiation and progression.

First, it is important to be clear about the object of explanation of the multistage theory. The multistage theory aims to explain *patterns of cancer incidence in populations*. Frank characterizes the project of his book as follows: "To understand cancer means to understand the genetic and environmental factors that determine the incidence curve. To learn about cancer, we study how genetic and environmental changes shift the incidence curve toward earlier or later stages" (Frank, 2007, p. 1). To be sure, this is not *all* it means to "understand cancer"; there are many different objects of explanation in cancer research, and explanatory success depends upon one's object. Carcinogenesis itself is a complex process that involves causal factors outside the cell; Soto and Sonnenschein characterize carcinogenesis as: "the *complex process whereby a mutation ends up forming the tissue dislocation typical of the carcinogenic lesion*." (Ibid., 1999, p. xi). In other words, how we characterize the object of explanation changes the relevant explanatory story. Soto and Sonnenschein (1999) argue at some length that cancer researchers ought to be concerned with a different question altogether than has been the focus in carcinogenesis: instead of focusing on specific mutations and their role in cancer, focus should turn to the role of the tissue microenvironment in cancer initiation and progression. Indeed, this is an active field of research, which has yielded important insights into cancer progression (see, e.g., Xu et al., 2009). However, in evaluating the explanatory success of a theory, one needs to be carefully attentive to the goal(s) of explanation and historical context.

### 3.1. Data integration

The multistage theory began by integrating data: collecting information about cancer incidence and generating a set of characteristic age-specific incidence curves for different cancers. As early as the 1920s, descriptive epidemiologists began to determine that different cancer types have characteristic curves of age-specific

incidence. For instance, breast cancer incidence gradually increases with age, peaking at about 65, and then much more gradually rising and eventually leveling off, through old age. On average, most women who get breast cancer do so between 50 and 75.[4]

What the authors of the multistage theory argued was most striking about these age-specific incidence curves was not differences (e.g., between different cancers), but similarities. Whether a cancer is common or rare, age-specific incidence curves have a similar shape; cancer incidence increases by and large as a power of age. That is, when one compares age of incidence across different cancers, one may see these curves as a pattern, such that it becomes possible to conceive a common (general) explanation. The fact that cancer incidence by and large increases as a power of age suggests a stepwise, cumulative process underpinning this outcome. One of the core insights driving the development of the multistage theory is that we can learn about *processes* of carcinogenesis by studying *patterns* of cancer incidence at the population level. While it seems simple, the generation of log–log plots of age-specific cancer incidence was one of the most important steps in getting researchers to conceive of different cancers as a product of a similar process. The data integration of these patterns constrained the development of models of cancer's dynamics.

Patterns of age-specific incidence were only one of several sources of data that are brought to bear in developing mathematical models of cancer dynamics. Over the past several decades, new data has been brought to bear on construction and refinement of the models, from a variety of sources: cell biology, molecular genetics, toxicology, developmental biology, epidemiology and clinical studies of patterns of inheritance of disease. The initial idea that cancer could be a product of mutations was at first prompted by observations of chromosomal abnormalities in tumor cells by some of the first cell biologists (Boveri, 1914), and some of the first work inducing cancer in experimental animals by some of the first toxicological studies (Charles & Luce-Clausen, 1942). Charles and Luce-Clausen developed one of the first quantitative models of multistage theory, based on observations of skin tumors on mice painted with benzopyrene.[5] It was also becoming clear in the 1940s that shifts in the curve of incidence to earlier age of onset were correlated with exposure to various environmental insults. Armitage and Doll (1954) argued that a series of multiple hits could yield patterns of cancer incidence. They recognized this possibility after studying how patterns of incidence in lung cancer shifted to earlier age with longer duration and more extensive use of tobacco. They suggested that for sporadic cancers, cancer incidence increased roughly by $t^{n-1}$, where t is age, and n is the number of rate-limiting steps. Fitting the data to the curve of acceleration of cancer incidence for a variety of different cancers, it appeared that at minimum, $n \approx 6$–7 events. This theory was later developed by Nordling (1953) and Stocks (1953), who generated log-log plots of incidence for different cancers, and used these data models of cancer incidence to argue that *different cancers had different numbers of steps required for onset*.

In 1971, Knudson generated a very precise, testable hypothesis: "the hypothesis is . . . that retinoblastoma is a cancer caused by two mutational events. In the dominantly inherited form, one mutation is inherited via the germinal cells and the second occurs in the somatic cells. In the nonhereditary form, both mutations occur in somatic cells" (Knudson, 1971, p. 820). Drawing upon patterns of familial incidence of retinoblastoma (siblings with and without the disease), and marking the patterns of age of onset of bilateral versus unilateral cancers, Knudson developed a mathematical model that represented cancer as a product of at least two mutations to a dominantly inherited gene. Knudson's prediction was borne out. In fact, *RB* (the retinoblastoma gene) was the first and prototypic tumor "suppressor" gene, a gene associated with regulation of the cell cycle or apoptosis.[6] Knudson developed this hypothesis on the bases of patterns of age of onset; the patterns of incidence constrained the available hypotheses "from the top down." By comparing and contrasting children that acquired the disease earlier versus later, and in two eyes versus one, he was able to hypothesize that the earlier onset and more devastating cases were likely due to the fact that these children needed to pass through fewer steps to get cancer than sporadic cases. Not only has Knudson's RB gene been identified, but its mechanism of action and role in the onset of retinoblastoma is well understood.

The case of retinoblastoma is unique, in that it is a childhood cancer with a strong heritable component. Most cancers take decades to develop, and thus involve the acquisition of a series of mutations. But, understanding this case was an important framework for further work. Moreover, the familial data was integrated with subsequent work on the rates and character of retinal development (Hethcote & Knudson, 1978), much like work on the character of tissue renewal in the colon has informed understanding of how carcinogenesis develops in that tissue (Cairns, 1975). The multistage theory of carcinogenesis provided a "unified" theory, drawing upon these data. More recently, detailed understanding of the complex networks of genes that control the cell cycle, and the reconstruction of "phylogenies" of cancer progression linking stages of cancer development to particular genetic changes has been linked with the idea that cancer involves the acquisition of a series of mutations, epigenetic, and chromosomal changes over time (Gerlinger, 2012). Particular mutations associated with colon cancer progression and their mechanisms of action in the cell have been identified (Fearon & Vogelstein, 1990), and the role of epigenetic changes such as hyper- and hypomethylation in gene expression (Hu et al., 2005), were further integrated into the theory.[7] In

[4] To be sure, there are many relevant causal explanations for the age of incidence curves, and their changes over time. e.g., the PSA test's popularity led to a spike in prostate cancer incidence; this may have been due to better detection, and not necessarily an increase in (deadly) cancers, as seems clear from the fact that the spike in incidence and earlier discovery did not lead to a decrease in mortality. Models of age-specific incidence for different cancers are available through SEER database, a US government run catalogue of nationwide incidence generated by descriptive epidemiologists, and patterns of incidence in different countries are compiled by IARC. Descriptive epidemiology is different from "analytic" epidemiology, in that it involves the description of patterns of incidence or mortality from different diseases, rather than attempting to discern causal associations.

[5] In particular, Charles and Luce assumed that the carcinogen causes a mutation rate $u$, and that $t$ is the time since onset of the treatment, the probability of two hits is $(ut)^2$. If painting affects $N$ cells, then $N(ut)^2$ cells are affected, and the time between the second genetic hit and the growth of a papilloma is $i$. Charles and Luce-Clauson concluded that the number of tumors per mouse after the time of first treatment is $N[u(t-i)]$.

[6] Tumor "suppressors" are often described as involved in keeping the "brakes" on proliferation of cells, whereas "oncogenes" are often described as "pressing the gas" on cell proliferation. Of course, no single gene is responsible for cell proliferation or quiescence. Many genes in combination, along with the tissue microenvironment, immune systems, tissue type, age, stage of development, blood supply, and methylation patterns all contribute to whether a cell divides or dies. However, there are some genes that do play an important role in whether a cell is likely to continue proliferating.

[7] There are a variety of histories of cancer research. Soto & Sonnenschein (1999), make a striking contrast with Weinberg (2008) and Angier (1988), and more recently, Mukherjee (2010). Cantor's (2008) collection is a comprehensive overview of issues surrounding control and prevention. Proctor's *Cancer Wars* (1996) is a compelling history of the politics behind the science of cancer causation. Morange (2011) in contrast, considers the relationship between biological theory and cancer research, focusing on the basic science more so than the political and social context. Perhaps needless to say, there is no "neutral" history—some will inevitably see my version as "triumphalist," i.e., as favoring the "winners." I certainly cannot make any claims to a comprehensive history here, but I can claim to draw upon a variety of sources; e.g., a balanced account of the history of cancer research can be found, e.g., in Morange (2011), and a brief but accessible and engaging history can be found in *Nature* (multiple authors), 2006, "Milestones in Cancer Research" (http://www.nature.com/milestones/milecancer/timeline.html).

other words, data from very different sources and scales of analysis have been brought to bear on the construction (and reconstruction) of the family of mathematical models of carcinogenesis. Each source both served to support the core idea behind the models, that cancer was the product of a process extended over time, limited by the accumulation of step-wise changes, but also often complicated the models, incorporating new information to the larger picture of cancer's dynamics.

In sum, the multi-stage dynamic model of cancer involved integrating data from a variety of different sources. All of the following played a role: cell biology (identification of chromosomal abnormalities in tumors, suggesting that mutations may be involved in cancer), toxicology (tests on animal models suggesting the two-hit theory), epidemiology (evidence that the curve of incidence could be shifted to earlier onset in smokers, suggesting that the rate could somehow be increased), and eventually, clinical medicine and studies of hereditary incidence of cancer (yielding Knudson's hypothesis) and molecular genetics (discovery of RB). Each of these independent sources of data informed the construction of the mathematical models of cancer's dynamics.

### 3.2. Explanatory integration

The practice of modeling cancer dynamics today is the culmination of the history of a unifying explanatory theory of carcinogenesis: the multistage model. Unifying theories often trade scope and explanatory power for detail. Modeling often involves idealization and simplification; for instance, one might assume (falsely) that the rate of some process is constant, or that a series of events occurs continuously, and that each event is independent of the others. By and large, it's easier to build a mathematical model when one makes such simplifying assumptions, and one often builds upon a prototype, or a model of the same or a similar process with similar dynamics. Modelers in the tradition of cancer dynamics deploy both simplification and appropriation. For instance, some of the first models built in the 1950s made a number of assumptions:

Mutations are the exclusive rate limiting events in cancer progression.

Cell mutations are a Poisson process, or stochastic process where each event is assumed to be independent of others, and occurring in a given time interval, and a tumor might occur after k such events (Nordling, 1953).

Cancer initiation is a chance event, takes place suddenly, and has a specified transition probability density function, or rate of change per unit time for each tissue type (Armitage & Doll, 1954)

Each tissue type has a specified induction period, constant for all initiation events in that tissue, but varying between tissues according to some distribution.

Some of the above assumptions were known or suspected to be false; others were merely hypotheses at the time they were proposed. Armitage and Doll (1954) were quite candid, for instance, in granting that different individuals' responses to the same environmental insult might have different outcomes, or that the rates of change in the same tissue type might vary by individual exposed. It was gradually becoming apparent that different tissue types, for instance, the colon crypt versus the cycles of growth and regression in breast and endometrial tissues, involved different rates of turnover of cells, and different rates of accumulation of mutations. Where appropriate, modelers would cite evidence (e.g., from cytology, genetics, incidence of leukemia in radiation exposed individuals, or the epidemiological data on smoking and cancer incidence), to support one or another assumption in their models. Good modelers are careful to be clear about when an assumption used to construct a model is deliberate simplification or simply false, and when it is a hypothesis supported by evidence. Unfortunately, what starts as deliberate simplification may often be confused with actual hypothesis and latter reified into theory.[8]

Nonetheless, these simplified models explained the age of incidence curves. If cancer is a rate-limited, multi-stage process, then, cancer incidence increases with age, and different numbers and types of events explain why there are different curves for different cancers. If cancer involves many steps, and these steps have a constant rate, we can understand cancer as the endpoint of the gradual accumulation of changes to cells and tissues over time, where variations between different curves and different ages of onset are explained either by inherited mutations, environmental insults, for different cancer types, or, different rate of turnover of cells in different tissues and different tissue architecture. The "explanation," in other words, is in terms of a general pattern or type of process, not the detailed causal mechanisms involved in setting the pace of these dynamics, in part because there is *no single* causal mechanism. To be sure, this explanation appeals only to a general pattern or type of process (Strevens "kernel"), not the detailed causal mechanisms involved in setting the pace of these dynamics, in part because there is *no single* causal mechanism. For any specific cancer, there must be a much more complex story. And, smoking can "shift" a curve of incidence earlier. So too can familial mutations, such as APC or BRCA I or II.

Thus, there is in a sense a "unified" theory of progression to cancer, even though cancers in different tissues often have very different molecular genetic profiles, as well as distinctive tissue micro-environments and immune environments. Initially, the theory black boxed this variation, as well as the detailed mechanistic and functional changes associated with particular mutations. Today, not only are the specific sequences of mutations better known, but how and why they affect cellular growth is much better known. However, the important point to note in this case is that the explanatory target in the case of cancer's dynamics is not a complete picture of oncogenesis in *all its detail*, but patterns of age-specific incidence.

Frank usefully analogizes the dynamic progression of cancer with the flight trajectory of an airplane:

> Molecular technology promises to reveal the biochemical changes of cancer. With that promise has also come and implicit assumption: one will understand cancer by enumerating the major biochemical changes involved in progression and the linkages of biochemical processes into networks that control birth and death. But enumerating parts and their connections is not enough.
>
> Think about a large airplane. If you were on that plane, the flight trajectory is what you would most care about. Could you predict the flight trajectory if you knew all of the individual control systems and their complex feedbacks? Probably not, because an inventory by itself does not provide all of the rates at which changes occur. Even with all of the rates for component processes, it would not be easy to work out the trajectory.
>
> One needs to link the parts to the outcome: how do particular changes in components shift the plane's trajectory? One ultimately assigns causality to parts by how changes in the parts affect changes in the outcome.

[8] For instance, in 1999, two researchers, Soto and Sonnenschein, wrote a book-length critique and history of what they took to be misleading assumptions and misconceptions of the somatic mutation theory, or SMT. Soto and Sonnenschein argued that the SMT should be replaced by TOFT, or "tissue organization field theory" (1999, 2000, 2005). Sonnenschein and Soto argue that the SMT inappropriately identifies the "default" state of cells as quiescence rather than proliferation, and locates the causal origin of cancer within the cell. Instead, they argue that cancer is an "emergent" product of failure in controls on cellular growth by the surrounding tissue. As others have argued (see, e.g., Bertolaso, 2009, 2011; Malaterre, 2007), it's not clear that these two theories are necessarily in competition.

> In a similar way . . . to understand a particular type of cancer, we must understand the forces that shape the age-incidence curve and the forces that shift the curve. . .
> Perhaps we should wait for all the molecular and cellular details, after which the nature of progression and the final outcome of incidence may be clear. Unfortunately, enumeration will not work. The full list of parts for our plane does not tell us how it flies . . . (Frank, 2007, pp. 309–311)

Frank's analogy is instructive; the point here is very simple. Molecular geneticists have devoted much (worthwhile) effort to understanding the basic biology underpinning progression. However, in attempting to understand cancer, we should be concerned with more than simply enumerating all the molecular, genetic, cell, and tissue changes involved. Understanding cancer involves also understanding general patterns, e.g., of incidence over time, and in inherited versus sporadic cases. Such understanding requires a general modeling strategy, which (ultimately) will be integrated with the biochemical processes controlling cell and tissue growth. Indeed, arguably, without the multistage model, there would not have been the theoretical framework in place to go out and seek particular genes involved in cancer.

Of course, in any *general* mathematical or computational model of a complex biological process, at least initially, one needs to black box some elements of the component causal process, and this is no less true in the case of modeling cancer's dynamics. Today, however, a variety of data constrains construction of models both from the "top down" (age of incidence data serves to delimit the number of discrete changes yielding cancer, e.g., mutation, epigenetic events), and, from the "bottom up" (e.g., with improved understanding of genetics, epigenetics, cellular and molecular bases of cancer).

What all the models of cancer dynamics share is that there is a cascade of rate-limited events that eventuate in tumor formation; but different cancers will exhibit different series of events yielding cancer. For instance, anti-apoptosis mutations can lead to chromosomal instability, followed by loss of heterozygosity; mutations in genes controlling cell cycle and mismatch repair are affected, as are epigenetic changes in methylation or chromatin structure, intercellular signaling, and the extracellular matrix. But this is just one possible sequence of events; each cancer is likely different, though there are some robust or common sequential patterns and processes (e.g., see, Spencer, 2006). In other words, constraint on allowable dynamic models of cancer is *also* "bottom up." This analysis of modeling of cancer's dynamics is compatible with Craver's (2007) mosaic, inter-level integration, and also, with Bechtel's (2010) account of combining intra-level causal relations and inter-level causal relations, or "top down" and "bottom up" explanations.

Cancer is a heterogeneous, and complex phenomena; while the multistage theory may account for or explain curves of cancer incidence for many cancers, it may not capture the complex causal dynamics underpinning each and every cancer (Weinberg, 2008). But, this is not altogether surprising. As modelers are well aware, if we waited upon (much less attempted to include) complete information, we would never build a model. Second, and perhaps more subtly, modeling is an exercise in simplification, in service of getting a (however preliminary) explanatory framework for some phenomenon. If we see all cancers as having in common a similar "dynamics," we have a tool for approaching new cancers and new questions about these cancers by drawing upon familiar models or representations. However imperfect these models, they serve a pragmatic, or heuristic function—namely, to permit us to identify relevant similarities and differences between different cancers, or, we can see how they depart or conform to this general characterization. Moreover, Knudson's model (1971) made a very specific prediction, which was not only borne out, but also led to one of the most well-understood genetic and developmental changes underpinning cancer.

## 4. Conclusions

It is by now a truism that at least one of the aims of science, if not the central aim, is to understand the causal structure of the world (Salmon, 1989). In medicine, the aim is also (if not primarily) to change the world—to prevent and cure disease. To change the world, we must know how to intervene on the world. One may intervene effectively with less than complete causal understanding. Many medical interventions were known to be effective without any clear (and sometimes false) understanding of why or how they were effective. One can understand *that* some drug is effective without understanding *why*. Moreover, one might have a *partial* understanding of why a drug is effective; one can know *that* a drug somehow intervenes on a certain metabolic, cellular, or gene regulatory pathway, but now exactly *how* it intervenes (which gene, enzyme, cytokine, etc. it intervenes upon).

Scientists give partial, incomplete, and how possibly explanations. The adequacy of an explanation (at least in practice) is often contingent upon its purposes. Clinicians and researchers may have different purposes, targets, and so different methodologies and standards of explanatory adequacy. As different disciplines focus attention on different explananda, so their standards of explanatory success will differ, and depend importantly on the relevant contrasts to the fact to be explained (Broadbent, 2011). In much of the philosophical literature on explanation, these hard won observations about scientific practice are often off screen. Scientific theories or explanations are measured against some ideal of completeness, whether this is an ideal explanatory text, or the world itself. In advancing a normative philosophy of science (e.g., one that advocates for more integrative research, or, alternatively, for "unified" theories), one must ask what function such norms serve: e.g., what is the benefit of integrative research? Why, if at all, ought "unifying" theories to be preferred?

Philosophers of science tried to discover or characterize the "logical structure" of scientific theories, or provide a unified account of shared goals among the sciences, when no such account could be had. A consensus seems to be building that the practice of science might better be characterized as a patchwork of more or less "integrated" activities. Unifying theories may play a preliminary role in framing a research program; integration is a distinctive goal, but the two are not necessarily at odds. A unified or unifying theory may start as a simplified, idealized starting point, a model that, over time, might accommodate a wider array of causal factors, or complex interactive effects. That is, incorporating a more integrative perspective, either via integration of data, concepts, methods, or instruments in service of solving specific problems or addressing particular anomalies, is not necessarily at odds with adopting models that treat distinctive phenomena as of a kind, or due to common causal patterns or processes.

Integration has become a term with a great deal of normative weight; calls for more integrative science seems to suppose that more "integrative" science will, perhaps, consider a wider array of evidence, provide a more holistic and complete picture of the world, and to top it all off, cure cancer. While it would be nice if a single, well-defined strategy could do all this, when one attends to the practice of science, what one is more likely to discover are a variety of partial, incomplete, and idealized models, explanations and hypotheses, targeted at different goals, which turn out to have greater or lesser scope, and which are more or less "integrative." Indeed, dividing the subject matter and focusing on one very specific phenomenon (e.g., age of incidence curves) can yield many useful insights. In other words, a "divide and conquer" approach,

focusing on a specific target of explanation, can be an effective strategy, provided that the divisions are not arbitrary, but informed by sound empirical evidence, and provided that models deployed are later understood to be just that: hypothetical, idealized, or simplified, not "true" models of the world in all its complexity. Integration may be a local goal, appropriate to some contexts and not others. Woodward (2010) argues that explanations in biology should be "proportional," or that, depending upon the explananda, or, facts to be explained, the appropriate level of analysis may require "black boxing" detailed component causes.

Early models of cancer's dynamics did just this; they "black boxed" both exogenous causes and component mechanisms underlying cancer initiation and progression. Later models filled in these boxes, with more detail about the molecular and genetic changes underpinning specific cancers. The object of explanation in cancer dynamics is patterns of cancer incidence at the population level. The "explanatory" integration in cancer dynamics consists thus in showing how different patterns of cancer incidence are similarly caused by a multistage processes.

Some might claim that scientists who black box or leave off screen all the complex causal interactions involved in producing some phenomena are not engaged in genuine "integration." However, attention to the practice of science suggests that integration can be partial, incomplete, or come in degrees. In other words, explanatory integration is not singular in type, or in quality. One can integrate data, methods, and explanations, to a greater or lesser degree, and the process of integration is often an iterated one, requiring that we circle back and reconfigure our understanding of both the object of explanation and the explanation itself. Paradigm change of the sort that involves a wholesale rejection of one "theory" in favor of another is rare; what is more common is that two research programs or subdisciplines discover mutual interest in a common problem or family of problems, which they may jointly solve. In understanding carcinogenesis, rather than the paradigm change from gene-centric vision that Soto and Sonnenschein (1999) hoped for, there may instead be a gradual accommodation and integration of the more diverse perspectives underway (see, e.g., Malaterre, 2007).

There are two sides to the "integration" story. Some see in integration the ideal of a complete and comprehensive explanation. However, we should beware of freighting integration with the weight of explanatory "completeness." Woodward's account of how choice of levels of explanation is determined is a useful reminder that scientific practice is often a matter of choosing one's battles carefully. Similarly, Wimsatt (1976, 2006) argues in his account of heuristics of modeling that reductionistic research strategies (while they have their limits) can be very effective at solving specific problems, as well as testing the limits of a reductive or simplified model. Modeling cancer dynamics is an example; by initially developing and testing simple models one can see patterns, which lead to successful predictions, as well as exceptions, and new models and explanations. One's choice of the object of explanation may involve simplifying or idealizing the component causal process, and focusing on one feature of the process of cancer progression to the exclusion of others. This anchoring process inevitably leaves some (lower level) mechanistic details outside of the explanation. However, such a strategy has arguably proven one of the more fruitful ones in science; particularly in sciences that are concerned to understand complex causal processes like cancer, choosing one's battles may be the only way.

Theories evolve. Today, as researchers in this field are well aware, many factors affect carcinogenesis; it is an enormously complex process, involving not only changes to genes, but also cell-to-cell signaling, changes in the functioning of gene networks, epigenetic[9] changes (hyper- and hypo-methylation), changes in the intercellular micro-environment, changes in tissue architecture, changes in interactions with the immune system, and much more besides. Some of these features can and have been integrated into the multi-stage model, but by no means all. All mathematical models involve by necessity idealization; no single mathematical model could represent all the "causes" of cancer at every relevant temporal and spatial scale. It is exactly for this reason that biologists focus on one scale of analysis, and develop simple, idealized models. Such idealizations are enormously effective in predicting, e.g., patterns of incidence, or, in discovering potential modes of intervention. The multi-stage model is "integrative," but it is also an attempt at unified account of what is ultimately a heterogeneous and enormously complex process.

Integration as a goal is often contrasted with the goals achieved by simplification and idealization often associated with unification; of course, the two goals are often in tension. As Cartwright (1980) has argued, comprehensive theories may sacrifice explanatory power and scope for causal detail. However, simple theories with the goal of unifying a diverse array of types of phenomena may also be the starting point for more integrative work, both via challenges to the theory, and, with integration of insights from diverse fields. It may turn out that the multistage model of cancer dynamics captures the underlying process of carcinogenesis only 90% or 70% or perhaps only 50% of cancers. In part for this reason, explanatory integration will often (and perhaps inevitably) be an iterated process, yielding partial, context-specific and hard-won explanations, which may either over time either become more integrative, by drawing upon a wider array of data from different sources, or, conversely, may be found to fail at identifying the most salient causal processes, or explaining as wide a scope of phenomena as might have been hoped.

As with any mathematical model of a complex system, in modeling cancer's dynamics, one runs aground of the problem of trading off generality with precision and descriptive adequacy. As in so much of biology, in cancer research, there are few if any general laws (Kincaid, 2008). Instead, one finds families of models, hypotheses, and generalizations of varying scope. In describing and explaining general patterns and processes, biologists often generate families of models (Downes, 1992; Lloyd, 1984), and such models are often simplified and idealized representations of the systems of interest. Perhaps especially in mathematical biology, there is often a trade-off between generality and realism (Levins, 1966; Weisberg, 2006, 2013). This is no less so in the case of modeling carcinogenesis.

## Acknowledgements

Many thanks to Ingo Brigandt, Alan Love, Alex Broadbent, Jim Tabery, Danielle Endres, Christopher Hunter Lean, Matt Haber, Rob Gehl, Marta Bertolaso, the participants at the Minnesota Workshop, and three anonymous reviewers for extensive comments. Thanks also for support from the Social Sciences and Humanities Research Council of Canada (Standard Research Grant 410-2008-0400 to Ingo Brigandt) for supporting the Minnesota Center for Philosophy of Science's Workshop on Integration in Contemporary Biology.

[9] Some biologists prefer to use the term "epigenetic" to refer all and only to heritable variations not encoded in the genome from one *organismal* generation to the next (Ptashne, 2007); here I use the term in the broader sense of non-genetic changes that are inherited from one *cell* generation to the next (see, e.g., Gilbert and Epel, 2008, for an explanation and examples of this use elsewhere).

## References


Achinstein, P. (1987). Scientific discovery and Maxwell's kinetic theory. *Philosophy of Science, 54*, 409–434.

Angier, N. (1988). *Natural Obsessions*. New York: Houghton Mifflin.

Armitage, P., & Doll, R. (1954). The age distribution of cancer and a multistage theory of carcinogenesis. *British Journal of Cancer, 8*, 1–12.

Balmain, A. (2001). Cancer genetics: From Boveri and Mendel to microarrays. *Nature Reviews Cancer, 1*, 77–82.

Bechtel, W. (Ed.). (1986). *Integrating scientific disciplines*. Dordrecht: Martinus Nihjoff Pub.

Bechtel, W. (1993). Integrating sciences by creating new disciplines: The case of cell biology. *Biology and Philosophy, 8*, 277–299.

Bechtel, W. (Ed.). (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (2010). The Downs and ups of mechanistic research: Circadian rhythm research as an exemplar. *Erkenntnis, 73*, 313–328.

Bechtel, W., & Hamilton, A. (2007). Reduction, integration, and the unity of science: Natural, behavioral, and social sciences and the humanities. In T. Kuipers (Ed.), *Philosophy of science: Focal issues (Volume 1 of the Handbook of the Philosophy of Science)*. New York: Elsevier.

Bechtel, W. (2013). From molecules to behavior and the clinic: Integration in chronobiology. *Studies in History and Philosophy of Biological and Biomedical, Sciences. 44*, 493–502.

Bertolaso, M. (2009). Towards an integrated view of neoplastic phenomena in cancer research. *History and Philosophy of the Life Sciences, 31*(1), 79–97.

Bertolaso, M. (2011). Hierarchies and causal relationships in interpretative models of the neoplastic process. *History and Philosophy of the Life Sciences, 33*(4), 515–537.

Bissell, M. J., & Hines, C. (2011). Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nature Medicine, 17*, 320–329.

Bissell, M., & Short, B. (2009). Mina Bissell: Context is everything. *Journal of Cell Biology, 185*(3), 374–375.

Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren [*Concerning the origin of Malignant Tumors*, H. Harris, Trans., 2008]. Cold Spring Harbor Laboratory Press, The company of Biologists, Ltd.

Brigandt, I. (2010). Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. *Erkenntnis, 73*(3), 295–311.

Brigandt, I., & Love, A. C. (2012). Reductionism in biology. In E. N. Zalta, (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2012 ed.), URL = <http://plato.stanford.edu/archives/summer2012/entries/reduction-biology/>.

Brigandt, I. (2013). Explanation in biology: Reduction, pluralism, and explanatory aims. *Science & Education, 22*, 69–91.

Brigandt, I. (this issue). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Brint, S., Turk-Bicakci, L., Proctor, K., & Murphy, S. P. (2009). Expanding the social frame of knowledge: Interdisciplinary degree-granting fields in American colleges and universities, 1975–2000. *Review of Higher Education, 32*(2), 155–183.

Broadbent, A. (2011). Inferring causation in epidemiology: Mechanisms, black boxes and contrasts. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences*. Oxford: Oxford University Press.

Cairns, J. (1975). Mutation selection and the natural history of cancer. *Nature, 255*, 197–200.

Cantor, D. (2008). *Cancer in the 20th century*. Johns Hopkins University Press.

Cartwright, N. (1980). The truth doesn't explain much. *American Philosophical Quarterly, 17*(2), 159–163.

Charles, D. R., & Luce-Clausen, E. M. (1942). The kinetics of papilloma formation in benzpyrene-treated mice. *Cancer research, 2*, 261–263.

Craver, C. F. (2005). Beyond reduction: Mechanisms, multifield integration and the unity of neuroscience. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 373–395.

Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Darden, L. (2005). Relations among fields: Mendelian, cytological and molecular mechanisms. *Studies in History Philosophy of Biological and Biomedical Science, 36*, 349–371.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of science, 43*, 44–64.

Downes, S. (1992). The importance of models in theorizing: A deflationary view. *Philosophy of Science, 1*, 142–153.

Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell, 61*, 759–767.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Philosophical Transactions of the Royal Society of Edinburgh, 52*, 343–399.

Frank, S. (2007). *Dynamics of cancer: Incidence, inheritance and evolution*. Princeton, NJ: Princeton University Press.

Frank, S., & Nowak, M. (2004). Problems of somatic mutation and cancer. *BioEssays, 26*, 291–299.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosohpy, LXX*(1), 5–19.

Galea, S., Riddle, M., & Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology, 39*, 97–106.

Gerlinger et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *NEJM, 336*(10), 883–892.

Gerson, E. (2013). Integration of specialties: an institutional and organizational view. *Studies in Biological and Biomedical Sciences., 44*, 515–524.

Gilbert, S. F., & Epel, D. (2008). *Ecological developmental biology*. Sinauer Associates.

Grantham, T. A. (2004). Conceptualizing the (dis)unity of science. *Philosophy of Science, 71*, 133–155.

Hethcote, H. W., & Knudson, A. G. (1978). Model for the incidence of embryonal cancers: Application to retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America, 75*, 2453–2457.

Hoos, A., Eggermont, M. M., Janetzki, S., Hodi, F. S., Ibrahim, R., Anderson, A., et al. (2011). Improved endpoints for cancer immunotherapy trials. *Journal of the National Cancer Institute, 102*(18), 1388–1397.

Hu, M., Yao, J., Cai, L., Bachman, K. E., van den Brule, F., Velculescu, V., et al. (2005). Distinct epigenetic changes in the stromal cells of breast cancers. *Nature Genetics, 37*, 899–905.

Kellert, S. H., Longino, H. E., & Waters, C. K. (2006). Introduction: The pluralist stance. In S. H. Kellert, H. E. Longino, & C. K. Waters (Eds.). *Scientific pluralism (Minnesota Studies in the Philosophy of Science)* (Vol. 9, pp. vii–xxvix). Minneapolis: University ofMinnesota Press.

Kessel, F., & Rosenfield, P. L. (2008). Toward transdisciplinary research: Historical and contemporary perspectives. *American Journal of Preventative Medicine, 35*(2S), S225–S234.

Kincaid, H. (2008). Do we need theory to study disease? Lessons from cancer research and their implications for mental illness. *Perspectives in Biology and Medicine, 51*(3), 367–378.

Kingsland, S. (1995). *Modeling nature: Episodes in the history of population ecology* (2nd ed.). University of Chicago Press.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48*, 507–531.

Kitcher, P. (1999). Unification as a regulative ideal. *Perspectives on Science, 7*, 337–348.

Knudson, A. G. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America, 68*, 820–823.

Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth*. Berkeley: University of California Press.

Leonelli, S. (this issue) Integrating data to acquire new knowledge: Three modes of integration in plant science. *Studies in Biological and Biomedical Sciences*.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist, 54*, 421–431.

Lewontin, R. (2000). What population geneticist know and how they know it. In Creath & Mainschein (Eds.), *Biology and epistemology*. Cambridge: Cambridge University Press.

Lloyd, E. A. (1984). A semantic approach to the structure of population genetics. *Philosophy of Science, 51*(2), 242–264.

Love, A. C. (2008a). Explaining evolutionary innovation and novelty: Criteria of explanatory adequacy and epistemological prerequisites. *Philosophy of Science, 75*, 874–886.

Love, A. C. (2008b). Explaining the ontogeny of form: Philosophical issues. In A. Plutynski & S. Sarkar (Eds.), *The Blackwell companion to philosophy of biology* (pp. 223–247). Malden: MA, Blackwell Publishers.

Love, A. C. (2010). Idealization in evolutionary developmental investigation: A tension between phenotypic plasticity and normal stages. *Philosophical Transactions of the Royal Society B, 365*, 679–690.

Malaterre, C. (2007). Organicism and reductionism in cancer research: Towards a systemic approach. *International Studies in the Philosophy of Science, 21*(1), 57–73.

Maull, N. (1977). Unifying science without reduction. *Studies in the History and Philosophy of the Sciences Part A, 9*, 143–162.

McMullin, E. (1979). Discussion review: Laudan's progress and its problems. *Philosophy of Science, 46*(4), 623–644.

Merlo, L. M., Pepper, J., Reid, B., & Maley, C. (2006). Cancer as an evolutionary and ecological process. *Nature Reviews Cancer, 6*, 924–935.

Mitchell, S. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Mitchell, S. and Dietrich, M. (2006). Integration without Unification: An Argument for Pluralism in the Biological Sciences. *The American Naturalist*, 168, S6. In integration across biological disciplines: A symposium organized by Mark A. McPeek (December 2006), pp. S73–S79.

Morange, M. (2011). *History of cancer research eLS*. Wiley Life Sciences.

Mukherjee, S. (2010). *The emperor of all maladies: A biography of cancer, simon and schuster*. New York.

Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace, and World.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2005). *Facilitating interdisciplinary research*. Washington, DC: National Academies Press.

NCI (2012). *Integrative cancer biology program: Centers for cancer systems biology*. <http://icbp.nci.nih.gov/> Accessed 15.03.12.

Nordling, C. O. (1953). A new theory on cancer-inducing mechanism. *British Journal of Cancer, 7*, 68–72.

Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science, 194*, 23–28.

O'Malley, M., & Soyer, O. (2011). The roles of integration in molecular systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 58–68.

O'Malley, M. A. (2013). When integration fails: Prokaryote phylogeny and the tree of life. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*, 551–562.

Plutynski, A. (2004). Explanation in classical population genetics. *Philosophy of Science Association Proceedings, 71*, 1204–1214.

Plutynski, A. (2014) *Explaining cancer: Philosophical issues in cancer causation and prevention*. Oxford: Oxford University Press. in press.

Proctor, R. (1996). *Cancer wars: How politics shapes what we do and don't know about cancer*. New York: Basic Books.

Ptashne, M. (2007). On the use of the word, 'epigenetic'. *Current Biology, 17*(7), R233–R236.

Salmon, W. (1989). Four Decades of Scientific Explanation. In Philip Kitcher & Wesley Salmon (Eds.). *Scientific explanation: Minnesota studies in the philosophy of science* (Vol. 13, pp. 3–219). Minneapolis: University of Minnesota Press.

Soto, A., & Sonnenschein, C. (1999). *The society of cells: Cancer and control of cell proliferation*. New York: Springer-Verlag.

Soto, A., & Sonnenschein, C. (2000). The somatic mutation theory of carcinogenesis: Why it should be dropped and replaced. *Molecular Carcinogenesis, 29*, 1–7.

Soto, A., & Sonnenschein, C. (2005). Emergentism as a default: Cancer as a problem of tissue organization. *Journal of Biosciences, 30*, 103–118.

Spencer, S., Gerety, R. A., Pienty, K. J., & Forrest, S. (2006). Modeling somatic evolution in tumorigenesis. *PLOS Computational Biology, 2*(8), 1. e108: 0001–0009.

Spitz, M., & Bondy, M. (2010). The evolving discipline of the molecular epidemiology of cancer. *Carcinogenesis, 31*(1), 127–134.

Stocks, P. (1953). A study of the age curve for cancer of the stomach in connection with a theory of the cancer producing mechanism. *British Journal of Cancer, 7*, 407–427.

Strevens (2004). The causal and unification accounts of explanation unified—causally. *Noûs, 38*, 154–176.

Strevens (2008). *Depth: An account of scientific explanation*. Cambridge, MA: Harvard University Press. 2008.

Weinberg, R. (2008). Using maths to tackle cancer. *Nature, 449*, 978–981.

Weisberg, M. (2006). Forty years of the strategy: Levins on model building and idealization. *Biology and Philosophy, 21*(5).

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford: Oxford University Press.

Williams, N. (2011). Arthritis and Nature's Joints. In J. K. Campbell (Ed.). *Carving Nature at its Joints—Topics in Contemporary Philosophy* (Vol. 8). Cambridge: MIT Press.

Wimsatt, W. (2006). Reductionism and its heuristics: Making methodological reductionism honest. *Synthese, 151*, 445–475.

Wimsatt, W. (1976). Reductive explanation: A functional account. In R. S. Cohen & A. Michalos (Eds.), *Proceedings of the 1974 meeting of the Philosophy of Science Association* (pp. 671–710). Dordrecht: D. Reidel.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy, 25*, 287–318.

Xu, R., Boudreau, A., & Bissell, M. J. (2009). Tissue architecture and function: Dynamic reciprocity via extra- and intra-cellular matrices. *Cancer Metastasis Reviews, 28*, 167–176.