Information *&* Method

Logic for Translational Science

Frank P. DeVita

Submitted in partial fulfillment
of the requirements for the degree
of Master of Arts
in the Graduate School of Arts & Sciences

Program in Biotechnology
Department of Biological Sciences

COLUMBIA UNIVERSITY

2018

Abstract

How do information, rationality, method, and knowledge dynamics shape what's possible with science? How do we make maximum contact with reality through scientific methods and reasoning? When can we assert scientific facts and confidently act on them? In this work, I approach these questions through a study of scientific rationality, information, and philosophical logic in context with some practical and theoretical challenges in contemporary life science research. The productive aim is to outline tactics for understanding the relations among scientific knowledge, information, rationality, and action that optimize decision making in scientific situations. There is a practical focus on translational biomedicine and molecular oncology because these fields aim to bridge gaps between theory and practice, and the dependence of rational decision making on knowledge relations is critical to their success. In my view, these and other precision scientific approaches may be better understood and executed by adopting pluralist and dynamic perspectives on science, information, and knowledge that, I argue, make better contact with reality than some existing knowledge-making strategies that aim to precipitate isolated scientific facts from interacting, evolving, and causally related bodies of information.

Contents

Figures

Tables

Acknowledgements

No single piece of writing can possibly capture or mention the full scope of ideas, people, and experiences that led to the development of this work. I am forever grateful first and foremost to my parents, and their open-mindedness to my academic endeavors and much more. Without them, none of this would be possible. I must also earnestly thank my professors, friends, colleagues, and extended family for their receptivity to the coarse versions of these ideas that made their way into conversations over the last few years.

I owe no less than my utmost gratitude to the Columbia University Department of Biological Sciences Program in Biotechnology for their enthusiastic support and openness to uncharted academic territory and creative research. If not for the expertise and guidance of Lili Yamasaki, Stuart Firestein, Robert Pollack, and many others, these views may have not had a fair chance to develop. I must also sincerely thank the Columbia University Department of Philosophy, particularly Robert Stalnaker and Taylor Carman for countless insights and arguments that pressed me to constantly sharpen my own thinking and writing. Thanks also to Arthur Kuflik and Alan Morrison for their ethical and legal expertise, Ann-Sophie Barwich for her views on pluralism and

scientific method, and Dimitris Anastassiou for his applications of information science to molecular biology.

Much of the insight that spawned the concept for this thesis would not have been possible without my participation in and exposure to translational at Memorial Sloan-Kettering Cancer Center and lectures at Herbert Irving Comprehensive Cancer Center. My professional experience with Ingo Mellinghoff, Alissa Thomas, Antonio Omuro, Lisa DeAngelis, and the Department of Neurology at the former, and meetings and discussion with Piero Dalerba at the latter are essential to this work on an inspirational level. Thanks also to Whitney Winter for her scientific wit, drive and ingenuity.

In addition to personal acknowledgements, I also want to make some conceptual ones. The first is to pragmatism. Throughout my development as a graduate student, I've found the lens of pragmatism to be one with high focusing power. When posed with the plurality of problems and challenges associated with different aspects of biotechnology, the importance of context, connection, and interaction emerge as pivotal relations to understand. This, in my view, makes philosophical externalism a useful position to take up. Through it's constantly shifting gestalt of third- and first-person analyses, one's theoretical orientation is put into view against that of others from inner *and* outer perspectives, raising interesting questions about what kind of information,

evidence, and facts carry what relative weight in scientific investigations, philosophical arguments, and many other contexts. This perspective helps—this writer at least—see how things hang together in the broadest sense and attempt to evolve knowledge about narrower problems.

Finally, this work owes itself to the intellectual open-mindedness Columbia University and the willingness of its departments and faculty to consider and embrace many different perspectives over any single outlook.

New York City

May 2018

To all those who listened, encouraged,
challenged, and guided.

Preface

These ideas developed through gaining close acquaintance with the engineering and informatics aspects of contemporary biotechnology, professional exposure to the challenges of translational "bench-to-bedside" biomedical research, and critical consideration of the logic and practice of science. Their core philosophical foundations developed over a series of seminars in the Departments of Biological Sciences and Philosophy at Columbia University between 2016 and 2018 covering perception, pluralism, and theory of knowledge. As they struck me, these talks made apparent theoretical and practical questions associated with entrenched patterns of scientific reasoning, and the their impact on the burgeoning era of precision medicine and its revolutionary aims:

> "Precision medicine endeavors to redefine our understanding of disease onset and progression, treatment response, and health outcomes through the more precise measurement of *potential contributors* — for example, molecular measurements as captured through DNA sequencing technologies or environmental exposures or other information…"
>
> (Precision Medicine Initiative 2015, my italics)

Deeper thought about this endeavor of precision medicine suggest several possible strategies for executing the approach, and the potential for a information-driven paradigm shift in biomedical science. In my view, this possibility hinges greatly on understanding how these "potential contributors"

identified with biotechnology may contribute to determining the significance of molecular data in context with existing knowledge and information, new discoveries, and the methods we use to determine and act on that data.

In what follows, I consider theoretical and practical aspects of scientific activity that aim to convert information into knowledge and action. I think the dynamics of this conversion can be clarified through the use of logic. My main analytical targets are some theoretical presuppositions of precision approaches that may affect its realization and expansion over time. My focus is to explore these foundational presuppositions and consider their implications for scientific theory and practice generally. Through connections to ideas from outside the natural sciences—namely from logic and the philosophies of science, perception, and language—I offer some alternative ways to reason about the issues I think these presuppositions suggest, and test these alternatives on the hard puzzle that malignant glioma presents to molecular oncology.

The presentation is (loosely) structured as a systems analysis over philosophy, theory, methodology, and application. It employs concepts and techniques that, in my view, motivate *scientific pluralism* as a useful attitude for dealing with the challenges of translating scientific information into effective action and analyzing science generally. Logical methods are imported as analytical tools for coping with information in ways that could be generalized

and differentially applied to problems that cut across multiple fields of inquiry. This importation motivates the discussion of theoretical and practical aspects of scientific pluralism throughout, and critical consideration of how scientific enterprise, progress, method, and knowledge could be seen differently.

In sum, this work applies non-traditional methods to problems raised by traditional ones to motivate alternative styles of reasoning that may better suit the complex problems that arise in contemporary scientific and theoretical research. These problems often require a plurality of techniques to solve because they involve many different forms of information, and this thesis attempts to develop a strategy for understanding what data *means*, and how we can reason about acting on it to enact the changes in the world we intend. Getting clear on these matters involves examining the nature of scientific activity and rationality generally, and understanding how it is we do, or *could do* science. From this perspective we may have a better chance at understanding which methods may best realize many different scientific aims.

1 Philosophy

The Information Problem

Science has always prioritized validity and soundness. However, it is possible that given the pace of data generation achieved by next-generation sequencing technology that science's classic cognitive schemas are growing increasingly hard to square with the contemporary research format of quickly evolving and highly specialized investigations that generate terabytes of information on a regular basis. (Illumina's NovaSeq platform can sequence up to 48 human genomes per run, see Illumina 2018) This situation is challenging for fields such as molecular oncology and others that aim to use empirical discoveries and information to determine practical strategies with high probabilities of success to cope with complicated, multivariate natural phenomena. This aim is virtuous, however its realization is complex from the molecular level up through the levels of information, decision, and action.

However complex, multivariate natural processes such as diseases can in principle be investigated and analyzed into actionable and scientifically interesting conclusions. With torrents of data generated by large next-generation sequencing projects, integration of molecular profiling into medical clinics, and

the rapid adoption of electronic health records, the data needed to make strong evidence-based conclusions about scientific direction and medical action are available, however the accompanying analytic methods are immature. At the same time, biology and biomedical research are quickly morphing from heavily empirical disciplines to fields that can benefit greatly from advances in information science and computational theory. Such advances could be immensely helpful in *wrangling* data (Kandel 2011) that may be important drivers of scientific discovery.

Current efforts in bioinformatics rely heavily on methods imported from computer science, and these methods have made analyses such as genome-wide association studies (GWAS) possible. These studies can drill down to the level of single nucleotide polymorphisms (SNPs) between individuals, and relate SNP status to traits such as eye color, or to conditions such as cardiovascular disease and dementia (Pickrell 2014). However, despite recent synergy between biology and computer science, there are many aspects of biological and biomedical data analysis that demonstrate the need for novel and intelligent approaches to data categorization, storage and manipulation.

In order to take advantage of analytic methods to produce actionable conclusions about any phenomenon, one needs a thorough understanding of it. Achieving this understanding not trivial, especially for complex phenomena

such as diseases, which entail a plurality of interacting processes at several levels of organization. Tracking these phenomena and their different aspects to render them intelligible to scientific analysis is an important step in maximizing the utility of computation in life science, biomedicine, biotechnology, and elsewhere. Part of taking this step diligently, in my view, entails getting clear on some implications, problems, and challenges that arise from having a vast landscape of information available for analysis.

Scientific Semantics

Given the explosion of sequencing and other data generated by clinics and laboratories since the Human Genome Project (International Human Genome Sequencing Consortium 2001), we have vast quantities of molecular information about which to theorize, but less of a grip on what kind of methods will render this information into maximally actionable conclusions—that is, into conclusions that hook onto the world such that we can readily change and reliably produce the effects we intend. In biomedicine, these effects are the treatment of disease and the improvement of health through biotechnological methods such as targeted therapy, immunotherapy, antibody therapy, and molecular profiling. These approaches and others are meant to provide scientists and practitioners

with the information needed to uncover causal mechanisms and exploit them, but a lack of systematic understanding about how exactly these general mechanisms relate to instantiated cases is a limiting factor in making revolutionary progress in major health concerns such as cancer, Alzheimer's disease, Parkinson's disease, multiple sclerosis and amyotrophic lateral sclerosis. In my view, the tension in this situation can be framed in terms of a classic distinction in linguistics and the philosophy of language: we have access to the *syntax*, or composition of these conditions at the molecular level as a result of sequencing technology, but much less of a grip on the *semantics*, or meaning and implication, of that data on theory and practice.

Molecular profiling has been espoused as a major step forward for precision medicine, with more breakthroughs promised to come:

> "We have already witnessed early successes of precision medicine. These include, for example, the development of targeted treatments for cancer and cystic fibrosis that are effective in patients who share an underlying causal genotype…In addition, new subtypes of disease are increasingly being defined through molecular profiling of affected tissues, an advance that is expected to lead to more focused design and testing of both therapeutic and preventative strategies — for example, treatments for specific subtypes of a disease, or behavioral interventions tailored to specific subgroups of the population."
>
> (Precision Medicine Initiative 2015)

Despite at least a decade of this general research approach existing as *translational* research, many of it's theoretical and practical aspects of are only coarsely defined. For instance, what do individual molecular profiles tell us about disease

mechanisms in particular versus general cases? Changes in health status? Quantified probabilities of treatment success? General answers to these types of questions are surprisingly elusive in practice compared to the clarity of the material facts molecular profiling can provide.

Through the identification of DNA sequences, molecular profiles suggest possible treatments based on individual genomic variation. With molecular profiling data, a clinician may, for instance, decide to use or avoid particular treatment strategies depending on the impact of a patient's unique genetic variants on molecular signaling cascades with druggable targets. Molecular profiles suggest causal stories about individual disease etiology and progression that should link up with causal mechanisms discovered in the lab. Thus, the union of these data sets in principle harbors facts about oncogenesis in individual (case-specific) and general (mechanistic) senses. By analysis then, one should be able to reasonably project, from present states of affairs—that is, *from the data*—what actions are warranted under particular circumstances. That is, relations between information and knowledge should tell us actionable and theoretically important things about natural phenomena grounded in their material reality such that one can select the best possible course of action given present empirical facts and background information.

The move from information to action is essential to precision medicine, however the approach seems less precise in practice, and more like a guessing game than a scientific paradigm, at least to this writer. In medical practice, there are relatively weak expectations that molecular profiling will reveal something beyond what drugs *might* be efficacious in *some* cases of disease for *some* individuals. Molecular profiles are more often used as tools for identifying and qualifying patients for clinical trials with genomic inclusion criteria rather than as maps connecting molecular information to rational action they should, in theory, be. The disposition characteristic of the active paradigm of "rational medicine" is to act on the conclusions of large clinical trials, and use molecular profiling results, if available, as supplementary aides. When clinical trial enrollment is not a factor, precision tactics tend to recede even further into the background of medical decision making.

Changing this entrenched disposition seems necessary for precision medicine to thrive. Some strategies to this end include the development of "basket trials" in which patients are given targeted therapy based on the presence of genomic variants in several different tissues of origin, however the method has raised more issues than it has solved—results have been arguably lackluster, and cannot get around many of the same road blocks of traditional strategies, such as tumor heterogeneity and misleading pre-clinical evidence that

doesn't translate from the lab to the clinic. (Mardis 2018) The results of a recently published basket trial of the EGFR inhibitor neratinib (Hyman et al. 2018) embody these issues, reporting that while the technique was helpful in identifying genomic variables of significance for disease activity and progression, therapeutic impact was significantly lower than that achieved by standard available therapies.

This situation exposes that biotechnology and molecular medicine have a *projection problem* under the precision paradigm—there is a lot of information that *should* clearly guide action and predict success, but it doesn't *necessarily*. The precision paradigm says to consider facts about *individual* genetics, environment, history and behavior in the context of a medical culture that generally prioritizes facts about *groups* studied in large clinical trials over such facts. This raises a logical problem for getting the precision project comfortably off the ground. Often, investigations that would furnish the data sets needed to reach high-confidence conclusions about the importance and influence of individual-level factors on health outcomes are studied as secondary, or worse, "exploratory" endpoints in clinical trials designed primarily for obtaining drug approvals—these are often unpublished, buried in appendices of scientific papers, or kept at the chest of private companies. This, in conjunction with open questions and some skepticism about the precision project (Simon 2008, Bayer & Galea 2015),

substantially reduces the theoretical and practical force of precision methods in the minds of clinicians in practice, however clear efforts are underway to change this disposition.

Research institutions and biotechnology companies such as Memorial Sloan-Kettering (Cheng DT et al. 2015) and Foundation Medicine (Frampton et al. 2013) are trying to render test results more accessible and actionable with detailed output reports, however those reports often come too late when the threat of mortality is high, are perceived as incapable of supplanting medical intuition, or are sometimes too ambiguous or vague to confidently act on without significant auxiliary motivations and confirmation. If precision medicine aims to achieve paradigm status, the knowledge strategy of squaring group (type, kind) level facts with individual (token) cases may be untenable given the plurality of component variables. This is not to say that clinical trial results should be abandoned or supplanted as instructive for biomedical decision making, but rather that increased epistemic and practical value of molecular profiles and bioinformatics generally is essential for the expansion of the precision paradigm. In other words, precision methods and information need to be *contextualized* existing methods and knowledge to truly become part of strategic inclinations and decision algorithms. To understand how this contextualization process could

get going, in my view, we have to get clear on the general relations that hold between information, knowledge, and action.

To this end, import of another concept from linguistics and the philosophy of language is useful: *pragmatics* is the study of linguistic acts and the context in which they are performed (Stalnaker 1970). I think the pragmatic approach can be applied in science in a similar way as the syntax/semantics distinction discussed above. If we consider scientific and medical actions from the *outside*—that is, in the practical and theoretical contexts in which they are performed—success rests on the way we *interpret* new and existing information against the background data, or *common ground* (Stalnaker 2014) connecting particular cases, theories, and evidence. The ability to act on information with high confidence, in my view, gears into our grasping the meaning and implications of that data in relation to existing knowledge, empirical facts, and new discoveries.

Since *information* plays a key role here, we need a definition: With Stalnaker (1998) I take information to be something "carryable, transmittable, and analyzable" and that,"one thing contains information about another if there are causal and counterfactual dependencies between the states of one and the states of the other." Relating forms of information is a knowledge-based theoretical task that calls for new ways to discover, explore, and understand the relations among data, facts, aims, and action, in addition to causal and

counterfactual dependencies and their effect on knowledge states. An approach that aims to pin down the implications of these relations should, in theory, *maximize contact with reality* (Chang 2012) and thus be theoretically and practically useful. That is, if we understand information and action with this level of relational precision in a way that is consistent with our theoretical orientations and practical aims, our knowledge states can become finer grained in the setting of expanding databases and the normative or disciplinary concerns.

Modern medical reasoning is already expected to continually adapt to perpetual discovery. It now regularly considers experimental conclusions in clinical situations, and clinical concerns can exert influence on experimental design. This dynamic is essential to translational research—practitioners and scientists must weigh the possible and necessary implications of stipulated trials and experiments against observations, case-specific histories, first and third-person accounts, new discoveries, institutional expectations, expert opinion, and community consensus to estimate probabilities of success and strategically make decisions. Scientific reasoning considered in this way has a decisively *analog* character—individuals must parse large quantities of different types of data with little guidance or systematic theory of how to sort, relate, and draw conclusions from mosaics of changing information. The *cognitive load* of this task is quickly outrunning human rational capacity, making methods for arriving at reliable

conclusions and making strong predictions increasingly desirable, and perhaps, necessary.

There are significant motivations for designing methods that can make crunching data more compatible with human cognitive ability. The ability to query relations among large bodies of medical literature, high numbers of patient cases, and molecular facts for theoretical and practical reasons would both clarify and sharpen decision-oriented tasks and uncover actionable patterns in data. Efforts are launched and underway to develop such tools in oncology by curating genomic information and clinical data (Cerami et al. 2012, Gao et al. 2013; see cbioportal.org), however their development and data curation processes can be slow and cumbersome. Genomic functional annotations are curated by hand, mostly by post-docs, and development does not typically include parallel development of decision making aides. However, well-funded efforts are underway to improve the quality and relevance of these clinical-biological knowledge databases (e.g. Chakravarty et al. 2017), and major technology companies (e.g. IBM, Google) are putting their analytics technologies to the test in innovative collaborative efforts to process and categorize biomedical data.

In parallel with these and other bioinformatics developments in the last decade, advancements in analytics, machine learning, and artificial intelligence have changed the way we interact with, process, access, think about, and behave

with respect to information. Data is constantly collected about human behavior for specific purposes, and analytical tools are able, from huge torrents of constantly expanding and changing data, to mine insights about innumerable business and industrial concerns. Crucially, machine learning technology can *understand* the data it is designed to analyze, and so can adapt to changing patterns in the data as they happen. AI systems thus *know* our wants, needs, preferences, and activities based on the data we generate by acting. Although *digital,* artificial intelligence and machine learning are modeled on our own patterns of thought and, coincidentally, the recursive observation-experiment loop that defines the scientific method. That is, there is a *cognitive consistency* between human and machine learning methods. We do, after all, call it artificial *intelligence*.

If we can unleash these digital analytic tools on disparate data sets in science and medicine, the possibilities of what could be achieved are staggering. Empirically oriented computation that leverages artificial intelligence and human thought patterns together should in principle cope with vast quantities of information and evidence whose synthesis exceeds human cognitive reach, and so opens up wider fields of possibility and implication that can be used to determine action. So much was already made a public spectacle in 2011 by IBM, when their question answering computer system Watson appeared on *Jeopardy!*

and handedly defeated two human champions. Importantly, Watson *weighed* his decisions based on top candidate responses deduced from the source data he was primed with prior to the competition (Figure 1).

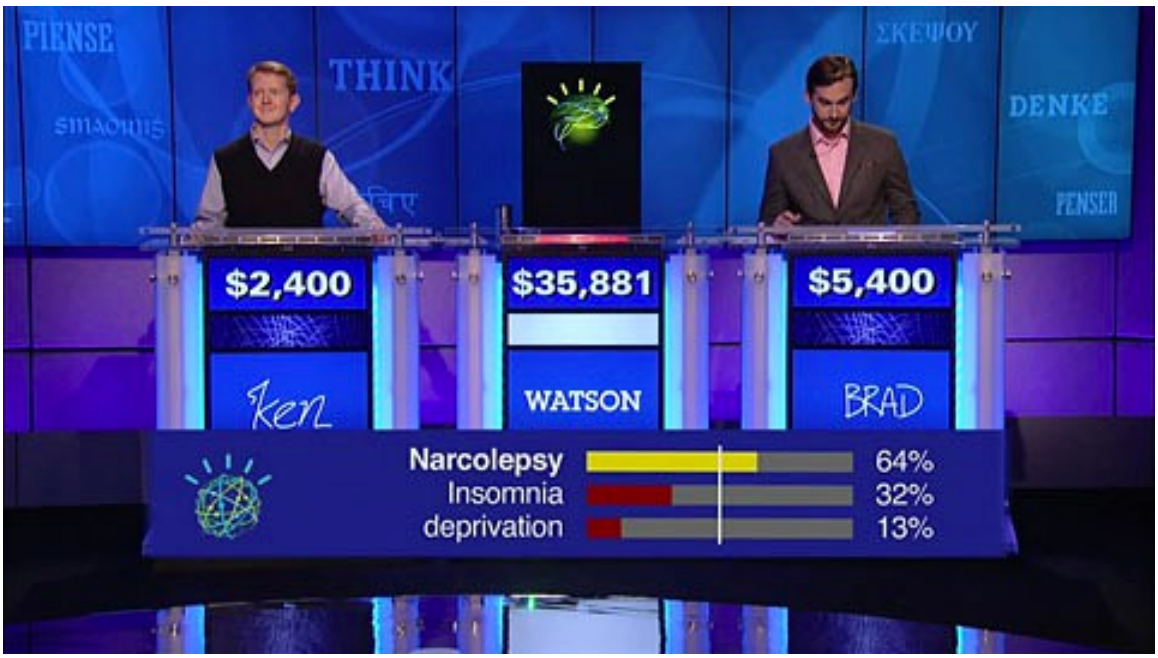

Figure 1: IBM's Watson weighing a decision on *Jeopardy!* On the nationally televised game show *Jeopardy!*, IBM's question answering computer system *Watson* (center) gained and held a commanding lead against human champions Ken Jennings (left) and Brad Rutter (right). Watson's weighing strategy, or "thought process," for response selection was shown at bottom.

Watson has since moved from primetime trivia to natural language processing and medical literature analysis. It has now "read" more scientific papers than any human—and has become a relational database and catalogue of medical syntax unrivaled by any other single source of information in science or elsewhere. Since Watson is a *language-based* system, it knows what experts have *said* about diseases, mechanisms, methods, and collected data in their published

work. What Watson doesn't know is what this expansive body of information *means* in human, consequential, causal, or counterfactual terms (cf. Searle 2011). Watson is a syntactical, deductive whiz, however we retain the semantic, inductive advantage by virtue of our cognitive ability to interpret information in conjunction with past experience, present knowledge, and context to act—we are *pragmatic* experts. A crucial point here, with connection to the philosophy of language, is that Watson and similar systems do not necessarily grasp the *pragmatic implicature* (Grice 1975) of conceptually related bodies of information. That is, the significance of particular relations against a shared body of contextual background thoughts, beliefs, information or other common ground elements. We deploy linguistic-computational algorithmic systems like Watson over linguistic information we've generated ourselves, and so such systems are only as effective as our queries to them are precise.

What matters to us are the implications of what we've said in the past about drugs, disease, genes, individuals, and groups in the context of present theoretical challenges and practical needs. Understanding the connections between this knowledge and data is the central aim of efforts such as IBM Watson Health, which promises to apply the cutting edge cognitive computing power of Watson to massive amounts of disparate health and scientific data, in principle surmounting many of the big challenges bioinformatics through

natural language processing (Chen et. al 2018). With high throughput sequencing and analytic technologies, we can process information in ways unrealizable in human minds and brains to sharpen inquiry and make more accurate predictions. This idea is not new (Goodman 1954), however, what the most versatile and reliable strategies for translational data analysis are, at least in biomedicine, undecided, and there is probably not a single approach that will cover all potential applications and research challenges. Getting clear on the nature of the cognitive and empirical frameworks we use in research should signal a path toward more general strategies that could relate disparate yet conceptually connected (and constantly expanding) bodies of information with reference to theoretical aims, practical motivations, existing knowledge, prior experience, and contextual common ground.

In the rest of this work, I aim to sketch an approach to inquiry and data analysis, with the intent of contouring a logic that can parse multiple different types of information with enough flexibility to adapt to situations in real time. It is precisely here where concepts and strategies from formal and philosophical logic are useful—crucially, they allow one to view and treat information differently and abstractly. This, in my view, permits richer analyses into possible implications of information and knowledge relations in interesting ways. Such techniques are potentially very useful for addressing challenges of translational

research generally, generating new perspectives on hard problems, and taking harder lines on new ones. The next section aims to get a better handle on the mechanics of scientific activity and thought in general so that we may use this understanding as a background for reasoning about the design of logical analytic methods.

2 Theory

Perception, Paradigms and Pluralism

The motivating thesis behind what follows is that if analytic methods are theoretically reflective of human patterns of thought and knowledge dynamics, they are that more likely to function as we expect them to, and thus guide effective action in practice. Phenomena are grasped in perception with a fundamental *figure-ground* organization. (Wertheimer 1912, Wageman 2012) They are in the world, yet perceived differently across individuals due to the unique perceptual *milieu* of individual perspectives on the world—that is, differences in cognitive background, knowledge, experience, training, and social embededness that structure perception (Merleau-Ponty 1945). The epistemological problem that arises from this phenomenological reality is posed by Kuhn (1962) in questioning the validity of an *observation language* directly related to sense impressions and analyzable in standard empirical ways:

> "…[M]odern psychological experimentation is rapidly proliferating phenomena with which that theory can scarcely deal. The duck-rabbit shows that two men with the same retinal impressions can see different things; the inverting lenses show that two men with different retinal impressions can see the same thing. Psychology supplies a great deal of other evidence to the same effect…"
>
> (Kuhn 1962, §X)

Here, Kuhn is generally puzzled about the relationship between perception and knowledge in scientific contexts. Putting Kuhn's point differently, we might say that perception has a *world-dependent content* and *individual-dependent rendering*. This distinction is useful for thinking about scientific activity and rationality if we understand empirical observation and classification as a specialized cases of this general perceptual dynamic (Figure 2).

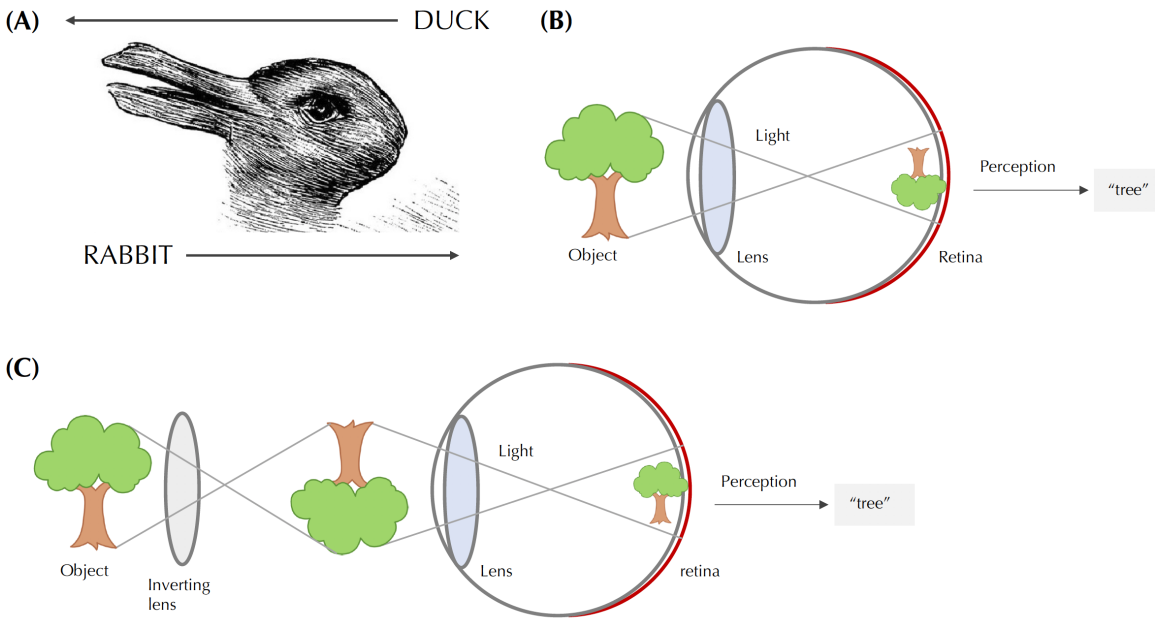

Figure 2: Epistemologically significant perceptual phenomena. **(A)** The duck rabbit is an unstable perceptual figure flipping between duck and rabbit. Which one *is* it? Can we say, both? **(B, C)** Normal perception of an object via retinal impression that results in the experience of a tree. An experimental condition with inverted goggles gives the eye a different (inverted) stimulus, and also produces the perception of a tree. *Sources:* **A:** public domain, **B & C**: author's drawings)

For example, a biologist and a physicist may study the same molecular phenomenon, yet explain it in different terms or concepts, and so too for scientists within the same discipline. Thus, observational language stands in tension with *theoretical language* that talks about unobservable properties, events or other scientific constructs, such as atoms or electrons (Blackwell 2004). If scientists can observe the same phenomenon yet classify it differently, the conditions for the possibility of this situation is epistemologically and practically interesting. It suggests that convergent observations with valid divergent classifications may be fundamental to science and scientific knowledge.

Complex natural phenomena such as perception, wave-particle duality, genomic regulation, health and disease, cognition, behavior, agency, and consciousness resist epistemologically *monistic* rational treatments, signaling a deficiency in such methods for these research objects. If the aim of science is to maximize informational contact with reality to learn as much as possible (Chang 2012), *scientific pluralism* is key to understanding how phenomena can be cubically observed and differentially classified. Under pluralism, like in Kuhn's musings, two individuals may have the same knowledge that *p* but validly differ in knowledge of how *p* is the case. One analysis is not necessarily better or worse than another, and they together condition a richer explanation of the phenomenon being observed. Cognition, for example, has biological,

neuroscientific, psychological, computational, and philosophical aspects, and this plurality of approaches taken together explains the phenomenon differently than any single approach or unified (monistic) or "interdisciplinary" theory. This dynamic between disciplines or approaches targeting the same phenomena maps onto the perceptual distinction between world-dependent content and individual-dependent rendering suggested above, and is consistent with Kuhn's worries about the foundations of scientific activity and knowledge. Scientific theories and observation are *contingent* on phenomena out in the world, and theories about them depend on one's field and purpose, intentionality, and the rules of logic.

Any science has a domain and a background on which observations are found and assertions made—that is, a set of *attractors* toward that bring its target phenomena, theoretical presuppositions, logical apparatus, and common knowledge into view. Sciences can share attractors, and attractors can play different roles in different fields with divergent theoretical and practical orientations. Multiple scientific fields and scientists within single fields can thus approach the different aspects of a phenomenon concurrently (Figure 3A), and reach divergent conclusions about it. In my view, this suggests that science is a networked *plurality* of theories and knowledge mutually concentrated on phenomena (Figure 3B), rather than a field of competing, independent or isolated

frameworks after the *singular set* of best explanations. Under this kind of scientific pluralism, theories and explanations may differ within and across disciplines, but together they constitute a set of contact points with reality through which information and knowledge can be organized.

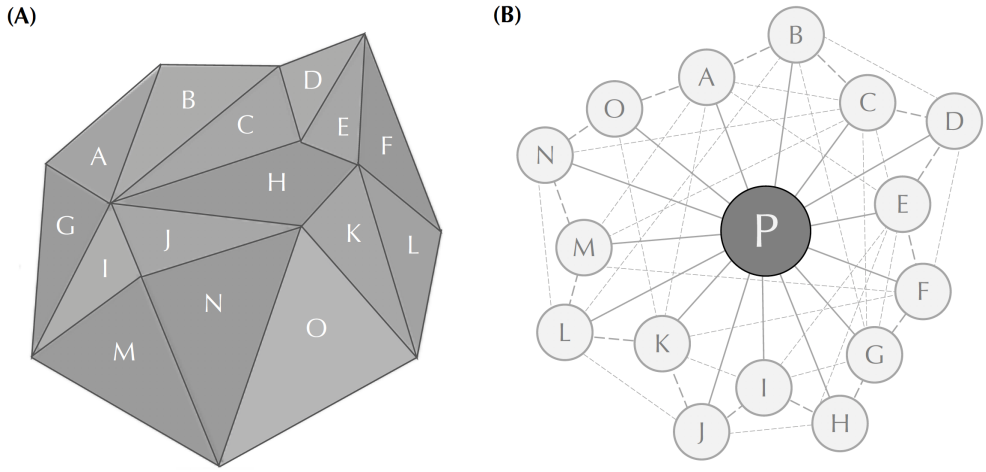

Figure 3: Phenomena and knowledge networks. (A) Pictorial representation of a phenomenon with multiple aspects (A-M) open for investigation. **(B)** Epistemic network constituting scientific knowledge of the phenomenon (P) through a plurality of approaches (A-M) targeting the phenomenon (lines) that are independent, yet connectable (dashes).

Pluralism challenges what we traditionally think about how science works, especially in contexts where disciplinary lines may blur, implications may cut across multiple domains, and anomalies don't lend themselves any particular established approach. That is, when observations made in scientific contexts are bound to admit to more than one rationally justified treatment (Chang 2012), and so are inherently open to multiple analyses. In this context, monist inclinations

toward theoretical unity and singular explanations lose their force to rich explanatory models that take multiple perspectives, theories, datasets, and differential conclusions into consideration. Furthermore, since pluralist models of explanation are rich in this way, they minimize logical dependence on any single set of evidence or method.

Scientific pluralism also challenges traditional conceptions of scientific explanation *without* undermining its force. Under pluralism, scientific explanation and knowledge can be treated like networks rather than a collection of isolated conclusions. Importantly, pluralism does not necessarily pursue fundamental theories, singular answers, or prioritization of one theoretical framework over another—it is motivated toward broader explanatory modeling. Thus, it also diverges from traditional views about scientific structure and progress, evidenced by its motivation of some unconventional and interesting accounts of science and knowledge, such as Dupré's (1981, 1993) *promiscuous realism* by which all scientific kinds are natural kinds and Fayerabend's (1975) *epistemological anarchism* by which all possible means of knowledge gathering should be pursued and developed concurrently. These more radical perspectives may be important understanding the nature, content, and advancement of science, especially as research becomes more complex, multidisciplinary, computational, and information-driven.

Since pluralism holds that no theory or discipline *necessarily* makes stronger contact with reality than any other, it is the network of theories and explanations about a phenomenon that best accounts for it. It follows from this structure that science can in principle sustain coexisting, independently developing lines of inquiry without issue, for no line of inquiry aims to replace or explain away others. Instead, explanations and theories from independent lines of inquiry all *refer* to the target phenomenon, constituting a knowledge network about it in which perspectives can interconnect and interact as a result of shared aims, but still exist, function, and develop independently to generate knowledge (cf. Figure 3B above).

If the structure of science is networked and dynamic as I've suggested, pluralism challenges rational inclination toward unified scientific theories and blind adherence to accepted scientific narratives. It also helps cope with the underdetermination of scientific theories by evidence (Quine 1951). Theories are said to be *underdetermined* when we do not have access to all the evidence of a phenomenon (Harding 1976). According to the *Quine-Duhem thesis,* no theoretical claims can be confirmed or denied in isolation from related and auxiliary hypotheses in the context of incomplete evidence or information. If this is the case in science (which I believe it is), scientific theories are *always* subject to revision in light of new evidence and are therefore evolving, malleable, dynamic

things. This point about the nature of science itself and challenges the explanatory sufficiency of any singular theory or monistic epistemological strategy for explaining natural phenomena and doing science. Without monism, the foundations of science and scientific progress diverge from traditional theories, and this departure, in my view, can help address the information problem and problems like it.

The Computational Paradigm

In *Structure,* Kuhn argues that science proceeds by the succession of *paradigms*, or sets of concepts and assumptions used by scientists to do the work they do. He defines paradigms as, "universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners." On Kuhn's view, evidence-based explanations become scientific orthodoxy until enough anomalous observations cause a *paradigm shift* of theoretical and methodological revision or replacement. After the shift, scientists *see the world differently* than under the previous paradigm.

Paradigm shifts are thus driven by community-wide changes in perception, action, method, and language. They are also driven by community-wide changes in method. When a new instrument or technique allows scientists

to approach problems differently, the form of experimental design, observation, and results shift accordingly and lead to new or revised explanations and theories. Classic examples of scientific paradigm shifts include Newtonian mechanics to quantum theory in physics, phlogiston to oxygen chemistry, behaviorism in psychology, and molecular biology in life science. Kuhn argues that science progresses in phases of paradigms, alternating between what he calls *normal science*, essentially to puzzle-solving, and *revolutionary science*, conducted in periods with mounting anomalies. In revolutionary periods, new paradigms are adopted, and normal science under the new paradigm continues until the next phase of revolutionary science and subsequent paradigm shift (Figure 4).

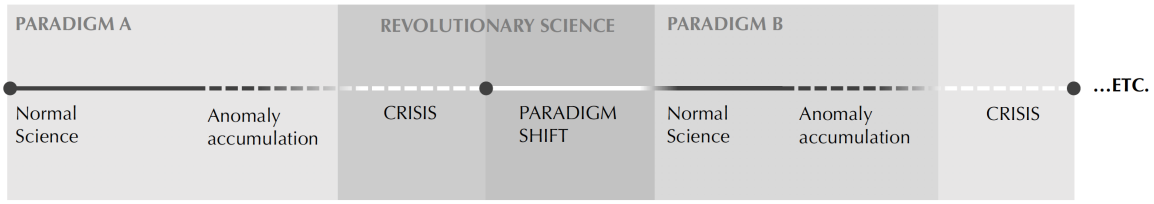

Figure 4: Kuhn's (1962) *Structure of Scientific Revolutions.* Kuhn argues for a historical, temporal, and phased conception of scientific progress characterized by periods of normal and revolutionary science that lead to recurrent paradigm shifts and new methods and perspectives developed to cope with gradually accumulating anomalies.

If there is anything resembling a paradigm shift happening now in life science, computational biology is a top candidate. Statistical and computational methods are being adapted for use in biology, yielding interesting results in

molecular science. Importantly, such methods help wrangle large quantities of data and uncover, for example, key molecular events in disease processes based on clustered expression data visualized as heat maps (Cheng et al. 2012). Interestingly, biological reasoning is not as aggressively or reflectively applied in the design of quantitative tools for analyzing biological data. Rather, existing computational methods are *applied* to biological problems and data rather than being developed *in context* with those problems in mind, construing computational biology as more of an applied computer science than a quantitative life science. This could be problematic, as it passes over the ground-floor integration of basic facts about living systems into things like disease modeling, bioinformatic analyses and other methods. This may leave algorithms unconstrained by common knowledge relations, and thus further from the actual conditions of the world they are meant to measure.

The main point here is that there is an important difference between applying statistical and computational methods to biomedical data and designing such methods from the ground up. Sociologically, this is demonstrated by an intellectual and cultural gap between the biological and computer sciences. Bioinformatics, computational biology, and systems biology bridge the two fields, however the individuals representing and teaching in these fields are often computer scientists or statisticians doing their work on a new type of

information, or biologists adding information scientists to their laboratory teams to analyze data. This may be because computational fields are advancing by solving biological problems and the converse is less true. Regardless, biological data continues to accumulate rapidly with the rising use of sequencing technology in medical and research centers worldwide, and methods are needed to wrangle it against existing knowledge to render the information actionable.

Kuhnian sociological worries aside, taming and learning from the data deluge by relating new information to what we've already discovered is necessary for scientific progress, however it may be defined. The integration of informatics strategies into the biological sciences opens the door to a new kind of biological thinking that I think counts as progress. Critically, the entities being studied are no longer only found on the laboratory bench—they are now also *represented* in data that can be analyzed to reach new conclusions. With the rapid development of next generation sequencing technologies over the past decade, research institutions are ramping up sequencing facilities and creating specialized research clusters dedicated to biological data storage, management and analysis (e.g., Columbia University, Stanford University, Memorial Sloan-Kettering Cancer Center, New York Genome Center). Simultaneously, new government initiatives have incentivized the transition to electronic medical records, creating a different, second information wave that occupies even more

space than sequencing data in the form of electronic medical records. Adding another dimension, wearable biometric devices represent another living body of information, and a method for data collection that ought to be used in conjunction with more traditional methods. On the sequencing side, there is an enormous need for the interpretation of stored information about DNA, RNA and proteins. On the health data side, there is an equal need for organized representation of information. Thirdly, sequencing data needs relating to clinical data to be rendered practically meaningful and informative for action. The varying forms of information involved in this kind of analysis warrant a logical method capable of dealing with those forms.

With the wealth of sequencing information now available there are important questions surrounding the best way to store and make this information available for use by biomedical and clinical communities. For example, researchers addressing problems in genomic analysis may be collecting data about patients with a particular disease at medical centers across the globe. The possibilities of using and analyzing this global body of information are scientifically exciting and medically important, however they may be undermined by incongruent methods and standards of data collection, storage and dissemination across institutions or research programs. These processes have not been generalized or standardized to drive discovery, and as such,

investigative efforts may remain individualized, diminishing the possibility for effective interrogation and manipulation of large, rich data sets containing clinical and molecular information. Moreover, it is an open question as to what the best analytic approaches may be. With much speculation around what sequencing data can means practically in different medical and scientific contexts, the instability of biomedical analytics could be approaching stagnation. I think the situation is less dire, yet still in need of critical examination—methods are computational, but they may not be optimizing action. Perhaps what is needed are refreshed rational directions and general consideration of the knowledge dynamics that condition confident action. One way toward these new directions is, I think, with pluralism (see Figure 5).

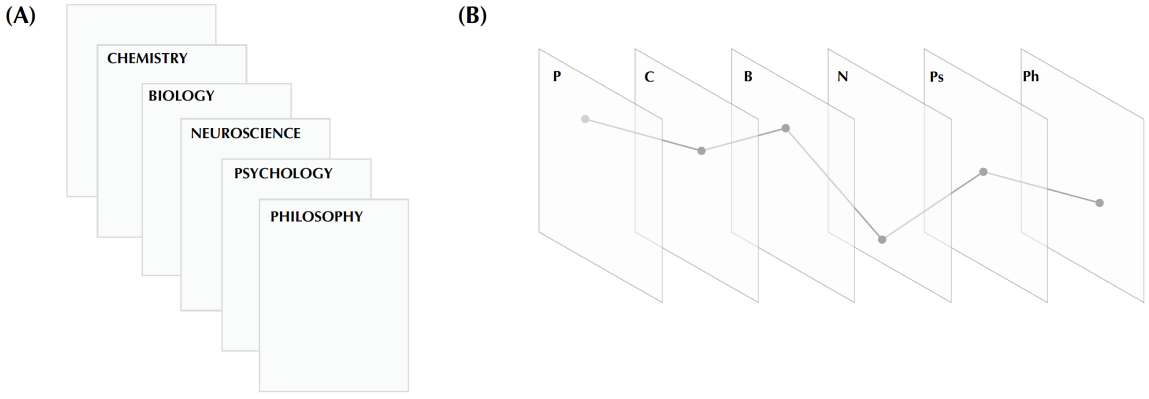

Figure 5: Pluralist theory of knowledge. (A) Scientific disciplines in isolation. Monistic knowledge strategies attempt to transpose information between domains without explanatory loss. **(B)** Pluralist knowledge strategy as an epistemological tool. Explanations link information across disciplines, or can be localized, e.g. within biomedicine with, for instance dimensions for molecular profiles, imaging results, treatments, and outcomes.

Above I've argued in favor of scientific pluralism as a philosophy of science that understands science's epistemic structure as a network with self-sustaining yet interconnected plurality of results and conclusions taken together constitute scientific knowledge. Like philosophical scientific pluralism, practical *pluralism about data* would accept that no type or piece of information should weighted, prioritized, privileged, or necessarily eliminated over another, and the aim is enriched, *dimensional* analysis—data sets should be maximized for breadth and cross-sectional analyses pursued to make new discoveries. This is possible with a pluralist epistemological strategy.

In the next section, I unpack some of the practical implications of this position by exploring how pluralism about classifying phenomena, collecting data, constructing databases, and analyzing information opens up new possibilities for scientific inquiry through logical computation. With this type of analytic method, information can be treated in new and creative ways since data can be *seen differently* depending on what relations are being queried. With this approach, a plurality of investigations with different practical aims could use shared, open data sets to test hypotheses and generate broader and deeper knowledge about natural phenomena. By this method, creative ideas about what bodies of information *could* demonstrate by relational analysis can drive scientific inquiry, observation, and classification, allowing investigators to work from the

phenomena up *and* the hypothesis down, with the limits of inquiry only constrained by the sharpness of our questions. With the rapid maturation of data science, analytics, and artificial intelligence, the possibilities realizable through such methods are dizzying. The next section takes this philosophical and theoretical hypothesis and tests it with molecular oncology to show how it might get going in practice. We'll turn to the mystifying puzzle of malignant glioma in molecular neuro-oncology. Molecular oncology faces difficult challenges under precision medicine due to its complex and multivariate subject matter. We'll start with a general assessment of the challenges in molecular oncology, then transition to the case of malignant glioma, and specifically its most deadly variant, *glioblastoma multiforme*.

3 Method & Application

Glioblastoma multiforme

Cancer is a disease with many forms that occur with some frequency across nearly all mammalian tissue types (National Cancer Institute 2018). Although the clinical phenomenology of the disease exhibits wide variation, the many different observed forms of the disease harbor overlapping genetic and molecular pathologies that cut across disease types, and many cancers share the same or similar molecular pathologies despite differing tissues of origin (Harris & McCormick 2010). In this sense, cancer exhibits a natural kind of *multiple realizability* (Putnam 1967, Fodor 1974) in which certain material properties or states are implicated across multiple related but distinct phenomena—different cancer types share genetic variations and gene expression dynamics, and the latter further depend on cell signaling, transcription factor action, and other molecular phenomena. Further, these phenomena all influence, and are influenced by, individual genetic background, exposures, and environmental conditions. On top of this complexity, individuals with the same cancers may also have different molecular profiles and respond differentially to treatments, making the delineation of clinical diagnostic categories and generalized care

algorithm construction difficult. This is acutely the case in neuro-oncology, in which malignant glioma presents some of medicine's most difficult diagnostic and therapeutic challenges.

Among brain tumors, malignant gliomas and their Stage IV variant glioblastoma multiforme (GBM) comprise 17,000 new cancer diagnoses per year and, although rare, these tumors are associated with, "dismal prognosis and poor quality of life": 1-year survival rates of 35.7%, and 5-year survival rates of 4.7%. With glioblastoma accounting for 82% of all malignant gliomas, neuro-oncology is faced with a deadly force for which no strong or reliably effective therapeutic strategy exists (Omuro & DeAngelis 2013). It is in puzzles like these where biotechnology and precision medicine are hoped to make a difference.

Malignant gliomas form as a result of a multistep process involving sequential and cumulative pathological genetic alterations resulting from internal and external forces. Adding again to the complexity, glioblastomas are histologically and genetically heterogeneous, and have at least 4 different subclasses with varying genomic alterations according to genome-wide expression studies: classical, mesenchymal, proneural and neural (Thomas et al. 2014, Table 1). Many of these alterations cut across GBM subclasses, and additional rare or novel mutations also appear in some cases.

Table 1: Genomic complexity of glioblastoma multiforme (GBM) subclasses.
Synthesized from Omuro & DeAngelis 2013, Thomas et. al 2014, and Lombardi & Assem 2017

	Classical	Mesenchymal	Proneural	Neural	G-CIMP
Chromosomal abnormalities	***Chr. 7*** amplification ***Chr. 10*** deletion				
Locus variations	***Ink4a/ARF*** locus deletion				
Single gene variations	***EGFR*** amplification/ point or *vIII* mutation ***TP53*** absence of mutation	***NF1*** mutation/deletion ***PDGFRA and nearby RTKs*** High amplification ***TP53*** mutation	***PDGFRA*** alteration ***IDH1*** *m*utation ***TP53*** mutation ***NF1*** deletion/ silencing ***PTEN*** point mutation	***EGFR*** over-expression	***IDH1*** alterations ***G-CIMP*** alterations
Expression dynamics		***CHI3L1, MET, and genes in TNF and NF–κB pathways*** High expression	***TNF*** super-family expression	***Neuronal markers*** expression/ over-expression	
Epigenetic variation					***MGMT*** hypermethylation

Classical, mesenchymal, proneural, neural, and G-CIMP tumor classes harbor different alterations that influence disease severity, progression, and treatment susceptibility or resistance. *Notes and abbreviations:* EGFR vIII mutation prevents binding of any known ligand; G-CIMP, glioma CpG island methylator phenotype

Therapeutic options for GBM are limited, but are expanding as a result of translating molecular discoveries made in the lab into clinically useful treatment strategies (Omuro & DeAngelis 2013). It has been discovered, for instance, that the most common site of mutation in primary glioblastoma is within the

promoter region of the telomerase reverse transcriptase gene, *TERT*, which increases RNA expression thought to be an important contributing factor to GBM development. These mutations are also found in lower grade gliomas, and in those cases, are mutually exclusive with mutations of *ATRX* and *IDH1* (Thomas et. al 2014). Complex molecular relations such as these are common in glioblastoma, thus complicating molecular analyses. Clinical complexity exists as well: non-cancerous syndromes may mimic malignant glioma on neuroimaging, requiring novel techniques for diagnosis (Omuro & DeAngelis 2013, Figure 6).

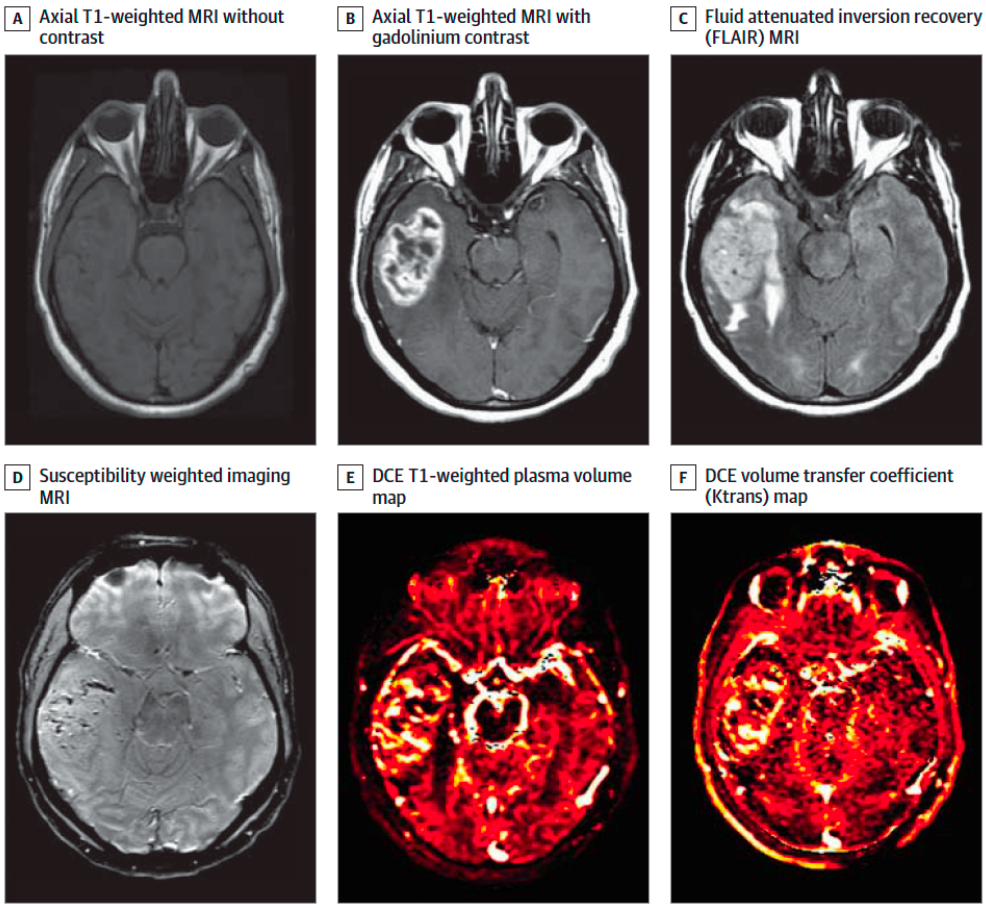

Figure 6: Neuroimaging studies of glioblastoma multiforme. Typical glioblastoma features on magnetic resonance imaging (MRI) studies used in the initial evaluation of a suspected brain tumor. Special techniques (B-C) must be used to see the tumor at all (A), and the extent of infiltration of tumor tissue into the healthy brain (from Omuro & DeAngelis 2013).

In these last few pages we've only scratched the surface of glioblastoma's complexity. Due to its complicated nature, GBM was the first solid tumor type to undergo comprehensive genomic, epigenetic, transcriptional, and proteomic analysis by The Cancer Genome Atlas (TCGA) effort. From this effort, it has been determined that over two-thirds of primary glioblastomas harbor mutations or amplifications of receptor tyrosine kinase (RTK) and growth factor receptor (GFR) signaling that lead to downstream activation in the PI3K-Akt-mTOR gene pathway affecting gene transcription, leading to cell survival (Cancer Genome Atlas Research Network 2008, Verhaak 2010). As a result the RTK (receptor tyrosine kinase) and GFR (growth factor receptor) pathways have been targeted as fruitful for therapeutic development, as drugs with similar mechanisms have been proven effective in other cancer types (Padfield et al 2015). Unfortunately for glioblastoma, trials testing the use of tyrosine kinase inhibitors of EGFR and PDGFR, mTOR, Akt, and PI3K have reported disappointing results (Omuro 2007). Aside from biological challenges such as the blood-brain barrier and tumor microenvironment, operationally most of these trials did not pre-select for patients with genomic variations susceptible to the targeted therapies they received (Thomas 2014). If precision medicine is to grow as a paradigm, opportunities like these must not be missed as a result of not knowing the clinical significance of particular variations in therapeutic contexts: better, systematic

screening methods were incorporated into trial designs, the "right" patient population could have been selected and the trial results could have turned out differently. As mentioned earlier, basket trials in oncology are attempting to seize opportunities to study variants in context, however early results indicate that there are still connections between pre-clinical evidence and medical action that remain to be understood.

Neuro-oncology is not alone in needing to cope with complex phenomena with a plurality of molecular events at their core. Biology has been coping with these types of phenomena for over a century. For instance, *pleiotropy,* the phenomenon of a single gene influencing multiple traits (Plate 1910), and *polygenic inheritance*, when one trait is influenced by multiple genes (East 1910 & Nillson-Ehle 1909 in Stearns 2010), bookend a spectrum that captures the worry. Put generally, and in context again with the philosophy of language, there are several kinds of *ambiguity* that arise in biological systems (Hodgkin 1998 via Empson 1930) at the genetic and molecular levels that make systematic analysis challenging. Pleiotropy for instance is a complex variable in the sense that it has a single input that generates a plurality of outputs clustered together through scientific and medical classification. There is a palpable lack of analytic approaches with which existing data and present knowledge are analyzed together to systematically *project* outcomes of possible events and courses of

action. These types of analyses have a high *translational utility* for both developing hypotheses and making decisions.

One road block in formalizing and categorizing biomedical data and maximizing translational utility lies in the complex causality that constitutes a biological fact. This much is illustrated by the analysis of glioblastoma above, but there are many other examples. For instance, polygenic traits such as height and skin color arise from the action of multiple genes in different body systems acting in a (presently unknown) temporal sequence. The underlying complexity of these traits is informatically dense such that achieving a full understanding of how, for example, an individual's height is determined, is no small feat. The situation is even more daunting in complex and poorly understood diseases such as glioblastoma, and other major contemporary health concerns such as Alzheimer's disease, multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), and Parkinson's disease, where molecular explanations are thought to be the rate-limiting step in developing effective treatments. These explanations, in my view, may be more easily attainable through the logical approach to genomics I'll now begin to sketch.

Logical Genomics

The application of formal and philosophical logic to information problems is one way to develop methods that can drive discovery about disease and other complex biological phenomena such as glioblastoma and pleiotropy. As mentioned above, over two-thirds of GBM cases have genomic alterations and corresponding signaling defects in receptor tyrosine kinase (RTK) and growth factor receptor (GFR) pathways. This broad state of molecular affairs can be represented by the conditional proposition $p \rightarrow q$ (if *p* then *q*), and specified for the context of neuro-oncology as:

(1) $mRTK \rightarrow GBM$

where mRTK = mutated receptor tyrosine kinase, GBM = glioblastoma multiforme

In natural language, this expresses the conditional proposition *if RTK is mutated, then GBM occurs. (*Note that in practice one would have to put this conditional relationship in context with the array of alterations in say, a mesenchymal GBM, and specify a specific RTK mutation, as a one or more may be mutated—the present example is simplified here for theoretical clarity.) If this proposition is *constrained* by the fact that over two-thirds of GBMs harbor RTK mutations, this

is a reasonable *logical representation* of some glioblastoma cases. The *truth-functional* outcomes of this two-term conditional are captured by the truth table in Figure 7, which exhausts all logical possibilities for the truth value of the conditional given all possible truth values (T or F) of its terms. All possible states of p and q—that is, the *truth conditions* of the conditional proposition—are captured in the table. As such, it is useful for reasoning about the influence of the of the component terms on the truth value of the conditional, in this case between GBM and RTK alterations.

p	q	(p → q)
F	F	**T**
F	T	**T**
T	F	**F**
T	T	**T**

Figure 7: Truth table for the conditional operator (⟶)
The conditional proposition p ⟶ q can be asserted as true, *except* if the antecedent (p) is true and the consequent (q) is false. This would create the fallacy of *denying the consequent.*

The output of row 1 for instance represents that for cases of the conditional in which *p* and *q* are both false, the implication is true—falsity implies falsity. Likewise, row 2 represents that when *p* is false and *q* is true, the implication is again true because *p* can be false and still imply *q*—we could say, for example that *if it rained then the ground will be wet,* but the ground could be wet for some other reason other than rain and this generalizes. Row 3 shows that when p is true and q is false, the conditional statement is false—if *q* follows from

p, then q is true when p is true and false when p is false (rows 1 and 4). Lastly, row 4 shows the obvious case that when p is true and q us true, the conditional proposition $p \longrightarrow q$ is true. Applying these conditional relationships to (1), we can reason that, for instance, there are *some* glioblastomas with RTK mutation (rows 3 and 4) and some without (rows 1 and 2), and that among RTK mutated brain tumors, there may be some that are glioblastoma and some that are not, and this matches clinical observation.

This way of reasoning shows that cases of disease can *in principle* be represented logically through the use of *propositional logic* (Table 2). The approach allows one to process the logical possibilities associated with particular cases and states of variables to represent facts and perform analyses. Another virtue of the approach is its *scalability*—one can imagine analyzing a large set of interconnected variables using various logical operators and representing complex cases, properties, and states-of-affairs with the terms p, q, r, s and so on with various operators to build structured database of cases for evaluating logical functions containing the terms of interest against one another to test hypotheses, derive conclusions and explore presently unanalyzed relationships, or different combinations of relations already known to be significant, such as point mutations in particular genes and particular kinds of epigenetic regulation. Through logic, these questions can be investigated as logical informatics puzzles.

Table 2: Usage and significance of logical operators.

Operator	Natural language	Definition	Use
¬	not	negation	¬p , “not P”
&	and	conjunction	p & q , “p and q”
∨	or	disjunction	p ∨ q , “p or q”
⟶	if…then	implication	p ⟶ q “p then q” , “p implies q”
⟷	if and only if	equivalence	p ⟷ q , “p iff q”
∃	there exists… for some… there is at least one…	existential quantifier	∃(p) p ⟶ q “there exists a p such that p implies q”
∀	for all… given any…	universal quantifier	∀(p) p ⟶ q “for any p, p implies Q”

Propositions are represented various operators to capture negation (¬), conjunction and disjunction (& , ∨), conditional relationships (⟶ , ⟷), and scope (∃, ∀). All can be used to represent information in abstract to explore relations and connections based on truth value.

In the general case of diseases driven by a combination of genetics, exposure, and environment, logical analysis enables many possible ways to relate different types of information that may shed light on causal mechanisms and presently unseen connections, as long as the data are structured such that they are quantified into logical forms. That is, such that data points represented by *p, q, r,* etc., make sense when substituted into propositions. Consider the following general case: disease *D* occurs when mutations occur in both genes *X* and *Y*, but not when there are mutations in only *X* or only *Y*. Let’s stipulate that we *know*

that a third mutation in gene Z conditionally amplifies the severity and progression of D in this context. We can represent *conditions* for D then with:

(2) $mX \,\&\, mY \longrightarrow D$ *(where mX, mY = mutated variants of genes X & Y)*

Further, we can represent severe (amplified) cases of D, say D^*, as:

(3) $(mX \,\&\, mY) \,\&\, mZ \longrightarrow D^*$

If (2) represents severe cases of disease with poor prognoses, we can then reason that since mZ confers increased severity, cases in which Z is not mutated likely have a better prognosis, and represent this with:

(4) $(mX \,\&\, mY) \,\&\, Z \longrightarrow D$

Similarly, we can reason out what cases are *not* occurrences of D or D^*, namely:

(5) $(mX \,\&\, Y) \,\&\, Z \longrightarrow \neg D$

(6) $(X \,\&\, mY) \,\&\, Z \longrightarrow \neg D$

(7) $(X \,\&\, Y) \,\&\, Z \longrightarrow \neg D$

(2) through (7) are simplified and generalized cases, but I think they sketch a viable *knowledge representation strategy* adaptable to different information contexts with little modification For instance, in a glioblastoma analysis we can define *X* as *EGFR, Z* as *MGMT* methylation, and so on for other variations.

If data is organized such that it is manipulable with logical operators as suggested above, the scope of possible relational analyses is enormous. As long as the facts of each case are accurately represented in logical form, the operators determine how variables like mutation status, disease severity, treatment resistance, disease progression, therapeutic success or failure get analyzed against one another. The approach can also maintain its integrity as more information and complexity from other domains such as imaging, vital signs, history, and symptoms are incorporated by adding more terms and operators. As the database acquires more information and dimension, more relations become available for exploration, and the existing information can be analyzed from different angles to litmus test primary and auxiliary hypotheses. These *angles of analysis* on the data in turn build and refine the knowledge network about targeted phenomena.

In order for this kind of logical analysis to be possible on a statistically significant scale in translational biomedicine specifically, sequencing information , current knowledge, and clinical outcomes must be stored together

(cf. Hawgood et al. 2015, Olson 2017). Raw sequencing data (e.g., ACTGGACCATA…) does not necessarily need to be represented in such databases for logical analysis to be fruitful, though it could be imported in analyses specifically targeting sequence variation effects—in those cases data points would be character strings that could be compared relationally to other character strings to evaluate similarity and difference. For biomedical applications however, genomic variation *states* (e.g., wild type, point-mutated, etc.) could be sufficient to compute, for example, predictions of response to targeted therapy given individual molecular profiles and known therapeutic targets. Analyses could become more exploratory by querying specifically for variants of unknown significance (VUS) that appear across cases of the same or related diseases, shedding light on their possible significance which could be followed up with laboratory experiments.

In the specific case of glioblastoma, we could use the growth factor pathways frequently activated in the disease state or other genes that have therapeutic targets to create a *data structure* that also captures possibly actionable or prognostic genomic variations such as *EGFR* amplification, and *IDH1* mutation, and *MGMT* promoter methylation (Thomas 2014, Benitez et. al. 2017, Lombardi & Assem 2017, cf. Table 1). This and similar information can be represented by a matrix of *M* rows and *N* columns such that cases and variables

are represented together (Table 3). This data structure represents an important first step in inquiry, as the decisions on what columns to use must be driven by current knowledge, or regulated by a hypothesis grounded in substantial empirical evidence, building in biological insight at ground floor.

Table 3: Data structure for storing and analyzing genomic information.

Patient	*EGFR* amplification	*IDH1* mutation	*MGMT* methylation
1			
2			
3			
4			
5			

Records of 5 patient hypothetical cases with 3 molecular variations important to treatment decisions. Truth values are assigned to represent facts in each case, e.g.“true” (T) for “EGFR application” (see Figure 4)

From here, we can begin populating the structure with the goal of relationally analyzing genomic alterations over different cases. Let’s stipulate that Patient 1’s glioblastoma is *EGFR* amplified, *IDH1* wild-type, and *MGMT* promoter methylated. According to present evidence (Thomas 2014, Benitez et. al. 2017, Lombardi & Assem 2017 & Cohen 2013) this tumor over-expresses EGFR (likely from a *vIII* point mutation that prevents ligand binding), is likely to progress quickly and acquire more mutations, and is more resistant to the

commonly used alkylating agent chemotherapy temozolomide. We can assign truth values to these variables and those of the other cases to prepare these cases for logical analysis (Table 4). Once prepared, the data can be interrogated logically, in a number of different ways depending on which vertical or horizontal relationships one may be interested in for various purposes, experiments, decisions, or actions.

Table 4: Truth value assignment to molecular variant status.

Patient	*EGFR* amplification	*IDH1* mutation	*MGMT* methylation
1	T	F	T
2	T	T	F
3	T	F	T
4	F	T	F
5	F	T	T

Molecular information captured as truth values. Truth value sequences (e.g. T-F-T) represent alteration load and can be evaluated as the terms (i.e. truth conditions) of a logical function.

With the data structured this way, we can analyze the data as truth table to uncover scientifically significant propositions and relationships across cases. These logical functions (see Stanford 2018 at web.stanford.edu.../truth-table-tool to explore possibilities), *if* they accurately represent causal relationships, or at least what we *believe* to be causal relationships based on the best available evidence, can be rendered actionable by logical analysis. For instance, Patient 1's

MGMT methylated, *EGFR* amplified, *IDH1* wild-type tumor profile can represented by:

(8) $EGFR \ \& \ \neg IDH1 \ \& MGMT$

Let's do a practical thought experiment: Say that since *IDH1* wild-type (unmutated) tumors like Patient 1's are more aggressive and acquire more genomic abnormalities faster, we want to query out from our database all *IDH* wild-type cases with *EGFR* amplification because we're interested in the mechanism of *IDH1* wild-type-accelerated progression and want to offer these patients EFGR targeted therapy as quickly as possible. Say we also want to know which patients in this set are also *MGMT* methylated (and thus less resistant to chemotherapy with temozolomide) so we can have a second line treatment plan for the (unfortunately likely) event that the EGFR targeting strategy fails. In this situation, the logical cases we're looking for are captured by the logical function $P\ (p,\ q,\ r)\ (\neg p\ \&\ q)\ \&\ r$ with p, q, and r standing for IDH1 status, EGFR amplification, and MGMT methylation, respectively. The three-term proposition (9) and its truth value table represent our molecular *query*, and illustrate the general approach for the logical treatment of molecular information.

(9) *P: (¬IDH1 & EGFR) & MGMT* …or…

p	q	r	((¬p ∧ q) ∧ r)
F	F	F	**F**
F	F	T	**F**
F	T	F	**F**
F	T	T	**T**
T	F	F	**F**
T	F	T	**F**
T	T	F	**F**
T	T	T	**F**

Figure 8: Logical treatment of molecular information. Truth table for *(¬p & q) & r* representing *IDH1* wild-type and *EGFR* amplification in the setting of *MGMT* promoter methylation.

The only case in which the function *(¬p & q) & r* (the generalized logical form of *P*) is when *p* is false, *q* is true, and *r* is true. That is, when an individual is not *IDH1* wild-type, is *EGFR* amplified), has *MGMT* promoter methylation that silences the oncogene. This logical discovery then suggests testable hypotheses. Looking back to our database (Table 4), if we tell an *algorithm* to pull all records with the value sequence that satisfies these truth conditions, it should return cases 1 and 3.

Directions for action then appear: In the laboratory, we can investigate any tumor tissue collected from surgeries these patients may have had that was banked for further study about *IDH1* wild-type oncogenesis in GBM. In the clinic, we can offer these patients EGFR targeted therapy, and be prepared with a second line plan for temozolomide plus radiation therapy (Stupp 2005) if and when the EGFR targeted approach fails. In a neuro-oncology clinic where a team of oncologists could evaluate and collect data on nearly 100 patients per day, this

kind of querying ability allows for the wrangling of rapidly expanding data sets that could easily grow to thousands of records in a few months. Add artificial intelligence to this situation and *live monitoring* can be deployed over the data to capture new records that meet the logical criteria of interest and anomalies.

In translational biomedical research, the goal is to discover insights about the connections between molecular biology and therapy, and we already know a lot about them, such as the causal mechanisms of acquired resistance to targeted therapy (Neil & Bivona 2017, Foo & Michor 2014) that can be incorporated into logical approaches. To illustrate this, let's stay with neuro-oncology and augment tour database to include clinical data in addition to the existing molecular data. In this *translational* database, decisions and outcomes are represented (Table 5).

Table 5: Data structure for translational analysis.

	Molecular Data			Clinical Data				
Patient	EGFR amp.	IDH1 mutation	MGMT meth.	Sx	Tx1	OC 1	Tx2	OC 2
1	T	F	T	T	rSx	PD	erlotinib	PD
2	T	T	F	T	lapatinib	PR	lapatinib	PR
3	T	F	T	T	lapatinib	PD	IT	PR
4	F	T	F	F	C/RT	PD	lapatinib	PD
5	F	T	T	T	gefitinib	PD	a-mTORi	SD

Logical approaches can relate data to explore hypotheses about the relations between molecular variation, treatment choice and sequence, and observed outcomes. *Abbreviations:* aVEGF, anti-vascular endothelial growth factor; C/RT, chemoradiation; IT, immunotherapy; mTORi, mammalian target of rapamycin inhibitor OC, outcome; PD, progressive disease; PR, partial response; rSx, repeat surgery; SD, stable disease; Sx, surgery; Tx, therapy.

As above, we can use logic to determine how we want to relate this data and bring different aspects of it into view. Say the database has hundreds of records. Then, a query can be built to investigate the effects and outcomes of treatment strategies and return records given particular molecular profiles and other relevant variables. For instance, we could query out cases with *IDH* wild-type and *EGFR* amplified tumors that are *MGMT* methylated that were surgically resected, then treated with lapatinib *L* with the function *Q*:

(10) $Q: [(\neg IDH1 \,\&\, EGFR) \;\&\; MGMT] \,\&\, (Sx \,\&\, L)$

Then we can tell the algorithm to return, for all cases with our stipulations in *Q*, the first outcome oc*1* of those cases:

(11) $Q: [(\neg IDH1 \,\&\, EGFR) \;\&\; MGMT] \,\&\, (Sx \,\&\, L) \rightarrow display: oc1$

Here, *Q* calls a list of cases in the target subset and their clinical outcomes. With this list, we can start to theorize about what kinds of treatment sequences are most effective for achieving partial responses given the outcome data, and use this information to inform *in vitro* and *in vivo* laboratory experiments as well as

clinical treatment strategies. Taking another step further, we can have the algorithm return only those cases where successive partial responses occurred:

(12) *Q: [(¬IDH1 & EGFR) & MGMT] & (Sx & L) ⟶*

display: ∀*(oc1, oc2) (oc1=PR & OC2=PR)*

Then, we can pull samples of the resected tissue form these cases for further laboratory analysis of the possible molecular mechanisms behind these (rare) positive results and investigate how the treatment sequences used may have elicited the partial responses. These and other analyses of the same form can enable a plurality of different investigations grounded in the manipulation of catalogued observations. Relational database tools that could implement this kind of logic exist (e.g., mySQL, see https://www.mysql.com), and it would be a fruitful exercise to determine if the capabilities of these tools can accommodate the approach sketched here.

I hope these experiments here have illustrated the value of formal and logic in data analysis. Logical approaches are general, flexible, and, in my view, allow for analyses more parallel with how we think, make decisions, and act on information. Again, the logic outlined here is a *sketch*, and further research is needed to work out the fine details and ensure consistency. There may also be ways to sharpen the method to align even more closely with natural cognitive

and knowledge dynamics. This is the attractor of the next and final section, which focuses on directions for further research, and aims at clarifying the cognitive, analytic, and empirical needs of biomedicine and other scientific fields in the setting of growing informatics challenges in contemporary research.

4 Future Directions

Knowledge and Information Dynamics

The reasons that motivate data collection in any science are pragmatic. They establish a *theoretical orientation* and an aim or aims toward and against which actions are taken and decisions are made. In modern life science and medicine, these decisions often circle around the discovery of molecular mechanisms in the lab that are exploitable in the clinic. To this end, clinical data including history, symptoms, exposures, behavior, mental status, physiology, demographics, disease activity, and treatment outcomes are tracked and recorded, and constitute the causal stories of individual cases. These data are meant to determine what clinical interventions are appropriate. In translational research, the primary aim is to *link* these data with molecular and other information about gene expression dynamics, genomic variation, cell signaling, metabolism, and other factors pertaining to the material nature of diseases. This strategy is meant to uncover causal mechanisms, drug targets, and other discoveries that can guide action. Thus, the translational move is to *synthesize* these molecular and clinical data to better understand their underlying causality when it is encountered in practice.

Complementary clinical and laboratory efforts create different types of related bodies of information: some with clinical information about particular individuals, cases and outcomes, others with causal information about disease etiology, progression, molecular properties, and some with a mix of both. This third type of data set is characteristically translational and can bridge the gap between medical and experimental science. However, as alluded to earlier, the forms of analysis best suited for interrogating the *union* of these data sets together are undecided, and it seems, at least to this writer, that efforts to determine these forms are not necessarily high priority research. Perhaps this is because such work, if it were to get things right, would entail matters beyond the immediate scope of traditional biomedical research projects and grant funding criteria, being work that requires information science, knowledge theory, formal analysis and a certain level of abstraction from classical research methodology.

Nevertheless, knowing what forms of analysis are best suited for such purposes would be useful even if their discovery traverses traditional boundaries. (This shouldn't be the case in my view. Scope definitions for scientific or other fields are a rich topic for discussion, and a full treatment of this issue is warranted elsewhere.) Here is precisely where pluralism about scientific research is important: Shouldn't research efforts with connections to a field's advancement, even those without the traditional contours of research typical in

that field, be embraced? This question is a poignant one in life science as it shifts and lends itself more and more to analytical methods traditionally found in fields such as information science, logic, knowledge theory, and operations research. *Applications* of insights from these fields to problems arising in the life sciences can perhaps begin to solve problems related to the information problem in biomedical research. The above study, I hope, has demonstrated the plausibility of this kind of *scientific pluralism* about research through an analysis of the information problem integrating insights and methods from these fields.

Bridging the information gap between theory and practice in science and elsewhere calls for nimble, adaptive logical frameworks and techniques that can enhance decision making through the active analysis of continually evolving, *dynamic* bodies of information. Dynamicism of this kind is an essential feature of modern data sets—they constantly expand and change over time as a result of new information. *Simpliciter,* the longer a target individual or phenomenon is followed, the more information we have about it. However, that information needs to be actively organized and packaged in an intelligible way *in context* with previously existing information and evidence in order to drive effective action. The data themselves do not necessarily constitute knowledge about the nature of the target or its future behavior, but they do in principle *i*ntegrate into broader epistemic and empirical contexts that contribute to decision making and

knowledge states. *Promoting* this information and evidence to knowledge that can reliably guide action requires knowing how it fits into different contexts of interest. How exactly this plays out in practice is less simple than simply stating the situation as a logical puzzle.

Methods that can convert information to actionable suggestions could have many different forms. The method outlined above is meant as a groundwork on which to build. By integrating probabilistic or Bayesian methods that estimate outcome events by evaluating not only sets of truth values within and over large numbers of cases, but of *probabilities* is one way forward. Depending on the truth value set of a particular case, one could determine the likelihood of an event's occurrence by the probabilistic composition of the case relative to others of the same form of the larger set. When patterns emerge, we might associate them with a *confidence threshold* to determine the uncertainty associated with various actions and act accordingly. However, action in reality is not as clear cut or simple as a computing "do *X*" by evaluating a finite set of binary truth values or estimating a probability. It is, nonetheless, a start.

Methods that take *modal value* and operators such as necessity (□), possibility (◇) and contingency (Kripke 1977, 2011) into consideration in rapidly or unpredictably evolving situations may be useful for constructing models with a higher degree of realism. Such methods could assign *rational credences*—that is,

*degrees of belief (*e.g., as values between 0 and 1)—to particular actions given sets of facts, predictions, possible events, or sets thereof and more closely mimic the phenomena of thought and action. In my view, these modal methods more accurately reflect the way we naturally make decisions: We rarely ground our decisions in explicit rational calculus, but rather in knowledge and belief relations that account for facts, what we know, what we believe, and what could be the case under various sets of conditions. Mimicking these patterns in a relational logic for scientific data analysis would harmonize our computational methods with our rational and cognitive dispositions, thus creating tools better suited for our human style of thinking, regardless of the particular project or task at hand. The relations we experience in thought and action are complex, but there is no reason in principle why they could not be incorporated into analytic methods and algorithm design. So much is already being done in applications of artificial intelligence to consumer behavior and economic data by private companies for maximizing profits through 'business intelligence,' so why not employ the same tactics to inform action in science, medicine, and analysis?

In order to realize such tools and methods in biomedicine, the way we think about biomedical and translational research may need general reconsideration. The characteristically analog way by which we presently approach biomedical problems may encounter serious difficulty dealing with

new, complex, interacting domains and dynamic information structures. In order to establish a logic to wrangle this data, we need a through understanding these structures—and our own cognitive ones. This is especially needed in fields like molecular oncology, where biotechnology is used to find aspects of disease invisible to traditional medical intuition. The missing ingredient is an intelligent analytics engine.

What's needed are methods for *skillfully coping* (Dreyfus 1973) with the particular information dynamics in biomedicine and translational research that, once understood, can allow for live monitoring, creative manipulation, and informed action in developing situations. In modern analytics, the tools for this task are called *dashboards.* They allow a user to collect, store, and manipulate information in terms of what they care most about knowing from sets of information. Dashboards aggregate data, execute live analysis, and create live reports that can be customized and tweaked according to various operational needs (Google 2018). Dashboards are most often used to analyze the performance of advertising and web traffic (Figure 9). However, before logical tactics can be successfully applied in this way, the information we want (or need) to analyze must be organized intelligibly to be maximally useful. An important factor in this maximization is the willingness of individuals, research centers, and bio- and other technology companies to work together, share data, and uncover insights.

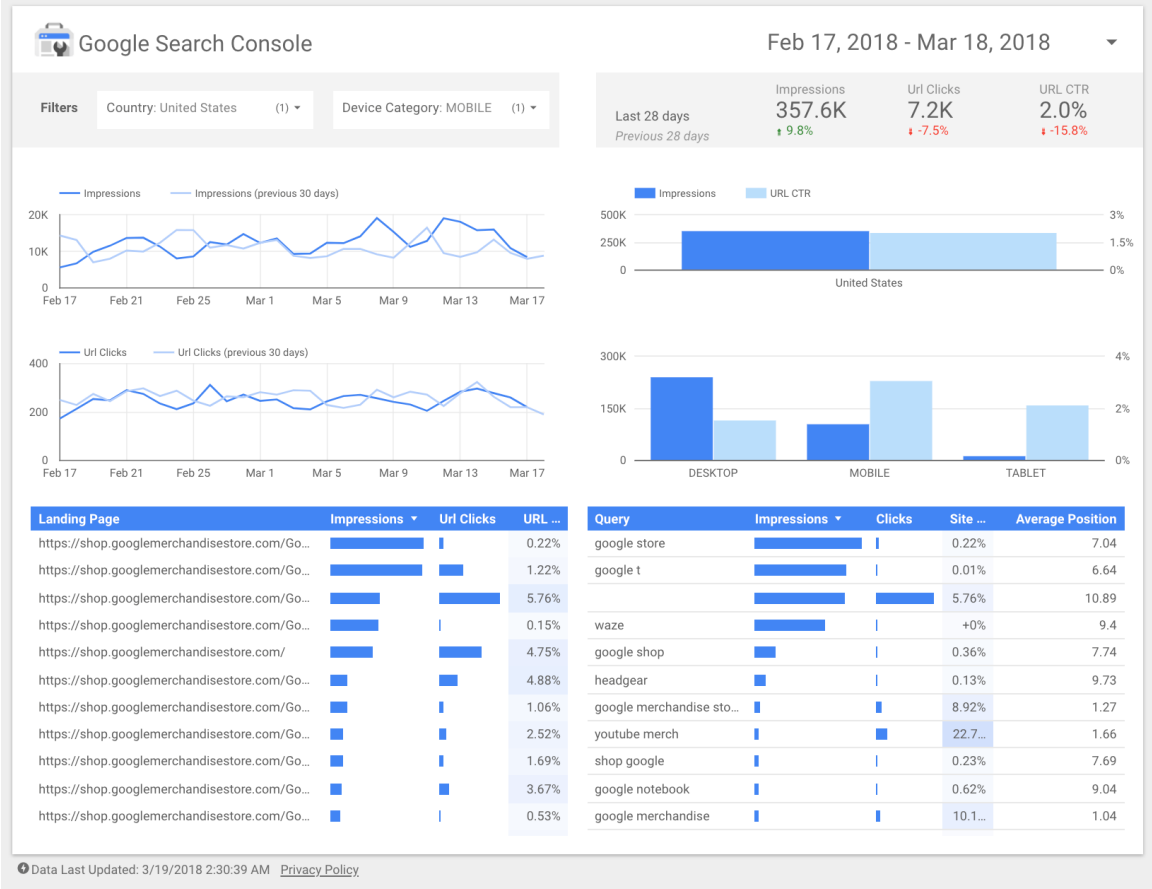

Figure 9: Google Analytics Dashboard for live analytics. Intelligent search activity in the United States shows search terms, impressions of results, URL clicks and result index position. The data is updated live, and visualizations are actively maintained. Trends can be put on different time scales and explored by frequency, activity level, or other user interests.

With cognitive computing platforms like IBM Watson and cloud-based data storage, monitoring manipulation software from Google, researchers and private entities alike are poised to turn ventures such as Watson Health into full-fledged, paradigm-defining research methods and practical tools. Applying these systems to scientific and biomedical information in combination with the logical methods sketched above could yield interesting results, enrich knowledge states,

sharpen inquiry, and enable more accurate projection from data. Logical genomics is one way to approach this organizational task that is nimble and generalizable enough to work in contexts inside and outside science and biomedicine. These tools can be used to generate powerful insights from vast amounts of information in real-time, and is thus one way to cope with information problems in science that, in my view, has equal theoretical and practical utility.

The general strategy advocated here is to logically and analytically approach how we interact with information, sharpen organizational principles, and emphasize the importance of changes in variables, terms, conditions, values, implications, and relative significance—that is, *dynamics* and *context*—in attaining states of knowledge and acting on it. This understanding is, in my view, sharply informative for the future development of analytic and intelligent systems that can account for these variables and the relations between them to suggest paths for action based on the best available evidence. Perhaps, through enriched understanding of these complex relations via these methods, computational analysis might mimic communication rather than calculation. With feedback loops, learning, and intelligent analysis improving our understanding of the natural world, we may also learn something about ourselves and our own rationality that swings open the doors of perception and knowledge to the kind

of contact and connectivity with the world promised by the information age. That is, the kind of understanding we could use to make this world, indeed, the best possible one (Leibniz 1710)—or, at least, to know if we're asking the best possible questions.

References

Bayer R & Galea S. Public Health in the Precision-Medicine Era. New England Journal of Medicine 373:499-501, 2015.

Bowman K. Integral Scientific Pluralism. *Journal of Integral Theory and Practice* 7(1), 54-66, 2012.

Blackwell Dictionary of Western Philosophy. Bunnin and You eds. *Observation language.*

Benitez JA, Ma J, D'Antonio M, Boyer A, Camargo MF, Zanca C, Kelly S, Khodadadi-Jamayran A, Jameson NM, Andersen M, Miletic H, Saberi S, Frazer K, Cavenee W & Furnari FB. PTEN regulates glioblastoma oncogenesis through chromatin-associated complexes of DAXX and histone H3.3. *Nature Communications* 8, article# 15223, 2017.

Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216):1061-1068, 2008.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery (*2), 401–4, 2012.

Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H1, Wang J, Rudolph JE, Yaeger R, Soumerai T, Nissan MH, Chang MT, Chandarlapaty S, Traina TA, Paik PK, Ho AL, Hantash FM, Grupe A, Baxi SS, Callahan MK, Snyder A, Chi P, Danila D, Gounder M, Harding JJ, Hellmann MD, Iyer G, Janjigian Y, Kaley T, Levine DA, Lowery M, Omuro A, Postow MA, Rathkopf D, Shoushtari AN, Shukla N, Voss M, Paraiso E, Zehir A, Berger MF, Taylor BS, Saltz LB, Riely GJ, Ladanyi M, Hyman DM, Baselga J, Sabbatini P, Solit DB, Schultz N. OncoKB: A Precision Oncology Knowledge Base. *JCO Precision Oncology* (1):1-16, 2017.

Chang, H. *Is Water H_2O? Evidence, Realism, and Pluralism.* Dordrecht New York: Springer, 2012.

Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, Dogan S, Borsu L, Hameed M, Nafa K, Arcila ME, Ladanyi M, Berger MF. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *The Journal of Molecular Diagnostics* (3):251-64, 2015.

Chen Y, Argentinis E., & Weber G. IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research. *Clinical Therapeutics* 38(4):688-701, 2016.

Cheng WY, Kandel JJ, Yamashiro DJ, Canoll P, Anastassiou D. A Multi-Cancer Mesenchymal Transition Gene Expression Signature Is Associated with Prolonged Time to Recurrence in Glioblastoma. *PLoS ONE* 7(4). e34705, 2012.

Cohen A, Holmen S & Colman H. IDH1 and IDH2 Mutations in Gliomas. *Current Neurology and Neuroscience Reports* 13(5): 345, 2013.

Dreyfus HL. *What Computers Can't Do: A Critique of Artificial Reason*. Part III. New York: Harper and Rowe. Toronto: Fitzhenry and Whiteside Limited, 1972.

Dupré J. Natural Kinds and Biological Taxa. *The Philosophical Review*, 90(1), 66-90, 1981.

Dupré J. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass: Harvard University Press, 1993.

East EM. A Mendelian Interpretation of Variation that is Apparently Continuous. *American Naturalist* 44: 65-82, 1910.

Empson W. *Seven Types of Ambiguity*. Chapters 1-8. Chatto and Windus, London, 1930. .

Fodor J. Special Sciences: Or the Disunity of Science as a Working Hypothesis. *Synthese* 28: 97–115, 1974.

Foo J. & Michor F. Evolution of acquired resistance to anti-cancer therapy. *Journal of Theoretical Biology* 21(355):10-20, 2014.

Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White J, Sanford EM, An P, Sun J, Juhn F, Brennan K, Iwanik K, Maillet A, Buell J, White E, Zhao M, Balasubramanian S, Terzic S, Richards T, Banning V, Garcia L, Mahoney K, Zwirko Z, Donahue A, Beltran H, Mosquera JM, Rubin MA, Dogan S, Hedvat CV, Berger MF, Pusztai L, Lechner M, Boshoff C, Jarosz M, Vietz C, Parker A, Miller VA, Ross JS, Curran J, Cronin MT, Stephens PJ, Lipson D, Yelensky R. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology* 31(11):1023-1031, 2013.

Harris TJR & McCormick F. The molecular pathology of cancer. *Nature Reviews Clinical Oncology* (7)5: 251-65. 2010.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* 6(269):pl1, 2016.

Goodman N. *Fact, Fiction, and Forecast.* 4th ed. Cambridge, MA: Harvard University Press, 1983 (first edition: 1954).

Google, Inc. (2018) *About Dashboards.* https://support.google.com/analytics/answer/1068216?hl=en&ref_topic=1068215 See also: https://www.google.com/analytics/analytics/features/. Accessed April 2018.

Grice HP. Logic and conversation, in P. Cole & J. Morgan (ed.), *Syntax and Semantics, 3: Speech Acts*, pp. 41–58, New York: Academic Press, 1975. Reprinted in H. P. Grice (ed.), *Studies in the Way of Words*, pp. 22–40, Cambridge, MA: Harvard University Press (1989).

Harding S. *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis.* Dordrecht-Holland Boston: D. Reidel Pub. Co., 1976.

Hawgood S., Hook-Barnard IG, O'Brien T, Yamamoto K. Precision medicine: Beyond the infection point. *Science Translational Medicine* (7)300:300ps17, 2015.

Hodgkin J. Seven types of pleiotropy. *International Journal of Developmental Biology* 42: 501-505, 1998.

Hyman DM, Piha-Paul SA, Won H1, Rodon J, Saura C, Shapiro GI, Juric D, Quinn D, Moreno V, Doger B, Mayer IA, Boni V, Calvo E, Loi S, Lockhart AC, Erinjeri JP, Scaltriti M, Ulaner GA, Patel J, Tang J, Beer H, Selcuklu SD, Hanrahan AJ, Bouvier N, Melcer M, Murali R, Schram AM, Smyth LM, Jhaveri K, Li BT, Drilon A, Harding JJ, Iyer G, Taylor BS, Berger MF, Cutler RE Jr, Xu F, Butturini A, Eli LD, Mann G, Farrell C, Lalani AS, Bryce RP, Arteaga CL, Meric-Bernstam F, Baselga J, Solit DB. HER kinase inhibition in patients with HER2- and HER3-mutant cancers. *Nature* 554:189-194, 2018.

Illumina, Inc. (2017) *High Throughput Sequencing Instruments Capable of Transforming Genomics to Improve Human Health at an Unprecedented Scale.* Press release. Jan. 9, 2017. https://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=2236383. Accessed March 2018.

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409:860–921, 2001.

Kripke S. *Naming and Necessity.* Blackwell Publishing Malden, MA, USA, 1980.

Kripke S. A Puzzle About Belief. *Philosophical Troubles: Collected Papers, Volume 1.* Oxford University Press, 2011.

Kandel S, Here J, Plaisant C, Kennedy J, van Ham F, Henry Riche N, Weaver C, Lee B, Brodbeck D, Buono P. Research Directions in Data Wrangling: Visualizations and Transformations for Usable and Credible Data. *Information Visualization Journal*, 10(4), 271-288, 2011.

Kuhn T & Hacking, I. *The Structure of Scientific Revolutions*. Chicago London: The University of Chicago Press, 2012 (first edition: 1962).

Leibniz GW. *Theodicy: Essays on the Goodness of God, the Freedom on Man and the Origin of Evil*. Edited by Austin Farrer and translated by E.M. Huggard. New Haven: Yale University Press, 1952 (first edition: 1710).

Lombardi M & Assem M. *Glioblastoma*. De Vleeschouwer S, editor. Brisbane (AU): Codon Publications, 2017

Mardis ER. Many mutations in one clinical-trial basket. *Nature* 554:173-174, 2018.

Merleau-Ponty M & Landes D. *Phenomenology of Perception*. Abingdon, Oxon, New York: Routledge, 2012 (first edition: 1945).

National Cancer Institute. SEER Training Modules. *Cancer Classification*. https://training.seer.cancer.gov/disease/categories/classification.html. Accessed April 2018.

National Institutes of Health. U.S. National Library of Medicine Genetics Home Reference. *What is the Precision Medicine Initiative?* https://ghr.nlm.nih.gov/primer/precisionmedicine/initiative. Reviewed April 2015, published February 27, 2018. Accessed March 2018.

Neil DS & Bivona TG. Resistance is futile: overcoming resistance to targeted therapies in lung adenocarcinoma. *npj Precision Oncology* (1)3, 2017.

Nilsson-Ehle NH. Kreuzungsuntersuchungen an Hafer und Weizen. *Lunds Universitets Arsskrift N.F.*, 5, 1909.

Olson MV. Precision medicine at the crossroads. *Human Genomics* 11:23, 2017.

Omuro AM & DeAngelis LM. Glioblastoma and Other Malignant Gliomas: A Clinical Review. *JAMA Neurology* 6(7): 1842-1850, 2013.

Omuro AM, Faivre S & Raymond E. Lessons learned in the development of targeted therapy for malignant gliomas. *Molecular Cancer Therapeutics* 310(17): 1909-1919, 2007.

Padfield E, Ellis EP, & Kurian KM. Current therapeutic advances targeting EGFR and EGFRvIII in glioblastoma. *Frontiers in Oncology* (5):5, 2015. .

Pickrell JK. Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *The American Journal of Human Genetics* 94, 559–573, 2014.

Plate L. *Genetics and Evolution* in Festschrift zum sechzigsten Geburtstag Richard Hertwigs. Fischer, Jena, Germany, 1910.

Precision Medicine Initiative. Working Group Report to the Advisory Committee to the Director, NIH. *The Precision Medicine Initiative Cohort Program —Building a Research Foundation for 21st Century Medicine*, 2015.

Putnam H. *Psychological Predicates* in W.H. Capitan and D.D. Merrill (eds.): *Art, Mind, and Religion.* Pittsburgh: University of Pittsburgh Press, 1967.

Quine WVO. (1951). Main Trends in Recent Philosophy: Two Dogmas of Empiricism. *The Philosophical Review*, 60(1), 20-43.

Searle J. (2011) Watson Doesn't Know It Won on 'Jeopardy!' *The Wall Street Journal.* New York, New York, February 23, 2011.

Simon R. Interpretation of Genomic Data: Questions and Answers. *Seminars in Hematology* 45(3): 196–204, 2008.

Stalnaker RC. Pragmatics. *Synthese* 2(1/2) Semantics of Natural Language, II, 272-289, 1970.

Stalnaker RC. *Context.* Oxford University Press, 2014.

Stanford University. Truth Table Generator. *CS103.* http://web.stanford.edu/class/cs103/tools/truth-table-tool/. Accessed April 2018.

Stearns FW. One Hundred Years of Pleiotropy: A Retrospective. *Genetics* 186(3): 767–773, 2010.

Stupp R, Mason WP, van den Bent MJ, Weller M, Fischer B, Taphoorn MJB, Belanger K, Brandes AA, Marosi C, Bogdhan U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Carincross G, Eisenhauer E, Miramanoff RO for the European Organization for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups and the National Cancer Institute of Canada Clinical Trials Group. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. New England Journal of Medicine 352(10):987-96, 2005

Thomas A, Brennan CW, DeAngelis, LM & Omuro AM. Emerging Therapies for Glioblastoma. *JAMA Neurology* 71(11):1437-1444, 2014.

Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN; Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17(1):98-110, 2010.

Wagemans J. A Century of Gestalt Psychology in Visual Perception: I. Perceptual Grouping and Figure–Ground Organization. *Psychological Bulletin,* 138(6), 1172-1217, 2012.

Wertheimer M. *On Perceived Motion and Figural Organization.* Ed. Lothar Spillmann. Cambridge, MA: MIT Press, 2012 (first edition: 1912).