Studies in History and Philosophy of Science 100 (2023) 107–115

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

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

Epistemic expression in the determination of biomolecular structure

Agnes Bolinska

Department of Philosophy, University of South Carolina, Columbia, SC 29208, USA

ARTICLE INFO

Keywords:
Scientific representation
Modeling
Efficiency
DNA
Protein structure

ABSTRACT

Scientific research is constrained by limited resources, so it is imperative that it be conducted efficiently. This paper introduces the notion of *epistemic expression*, a kind of representation that expedites the solution of research problems. Epistemic expressions are representations that (i) contain information in a way that enables more reliable information to place the most stringent constraints on possible solutions and (ii) make new information readily extractible by biasing the search through that space. I illustrate these conditions using historical and contemporary examples of biomolecular structure determination. Then, I argue that the notion of epistemic expression parts ways with pragmatic accounts of scientific representation and an understanding of models as artifacts, neither of which require models to accurately represent. Explicating epistemic expression thus fills a gap in our understanding of scientific practice, extending Morrison and Morgan's (1999) conception of models as investigative instruments.

1. Introduction

Certain representational devices help scientists learn more about the things they investigate than they might have been able to otherwise. For instance, Francis Crick and James Watson's determination of the DNA structure was facilitated by building physical models from precisely scaled pieces representing components of the molecule. It's not that, without molecular models, it was impossible to determine the DNA structure. Rosalind Franklin worked on the same problem without building models, and there is reason to think that she might have been the one to solve it (Maddox, 2002). It's just that these representations made the information Crick and Watson sought readily available; they were efficient; they saved time. Contemporary structural biology, too, depends crucially upon how available information is represented. A technique called integrative structure modeling, implemented in computer software, facilitates the representation of information in ways that enable the determination of biomolecular structures orders of magnitude larger and more complex than DNA (Rout & Sali, 2019; Sali, 2021).

In science, like in many contexts, research is constrained by limited resources, including time, funding, and materials. Scientists are motivated by prestige, and their careers depend upon how much they can publish, how much grant money they can attract, and how much influence their work garners. Pragmatic and sociological constraints like these are a reality that shapes scientific practice. There is no such thing as science conducted in some ideal world free from them. What's more, in many areas, such as biomedicine and climate science, research has significant consequences for the good of humanity. It is imperative, therefore, to conduct research as *efficiently* as possible—that is, in a manner that makes best use of limited resources.

The pernicious effects of sociological realities like the pressure to publish can affect the quality of scientific work by, for instance, leading to fewer reproducible results (Heesen, 2018). But this does not imply that we should avoid maximizing efficiency, lest it lead to a sacrifice in quality. Rather, we can understand efficiency in terms of the speed at which *high quality* research—research conducted conscientiously, without cutting corners—is produced. Given that efficiency is, as a matter of fact, important, how can it be maximized?

This paper is concerned with how the representations scientists use can play a central role in research heuristics that are efficient in this sense. To that end, let us call representations that expedite their users' ability to gather information about aspects of their target systems *epistemic expressions*. The verb 'to express' has several connotations. It can mean to convey, or to manifest, or to reveal, as when we express a particular feeling through body language or facial *expressions*. It can mean to represent, as when an equation expresses a mathematical function. Its root is the Latin *exprimere*, to press or force out, and it can have this quite literal meaning, too, as when we express the juice from an orange. As an adjective, 'express' can mean high-speed or direct, as in express post that delivers letters expediently or the express train that makes no stops between origin and destination.

The notion of epistemic expression draws on each of these meanings:

> Epistemic expressions are representations that (i) contain information about the things they represent and (ii) make new information readily

E-mail address: bolinska@mailbox.sc.edu.

https://doi.org/10.1016/j.shpsa.2023.05.009
Received 15 February 2022; Received in revised form 15 April 2023; Accepted 29 May 2023

extractible. They are thereby direct routes to solving research problems.

In addition to discussing epistemic expression *qua* kind of representation that facilitates efficient research heuristics, I will also consider it as the activity in which epistemic expressions are used.

The aim of this paper is to describe what makes epistemic expression possible by elaborating upon conditions (i) and (ii). I begin in Section 2 by characterizing the problem of biomolecular structure determination as an iterative reduction of a space of possible structures. Then, in Section 3, I argue that epistemic expressions expedite this process by (i) containing information in a way that renders the degree to which information constrains possible solutions proportional to its reliability and (ii) biasing the search through the possibility space toward the correct solution, making information sought by investigators readily available. In Section 4, I situate the notion of epistemic expression against the backdrop of two related conceptualizations of scientific models: as scientific representations understood pragmatically and as artifacts. I argue that although epistemic expressions share some of the features commonly attributed to scientific representations and models-as-artifacts, accounts of the latter notions do not have the resources to tell us how certain representations enhance research efficiency. The account of epistemic expression I offer, therefore, is a necessary supplement, illuminating an important aspect of modeling practice. I conclude in Section 5 by showing that thinking about epistemic expression can enhance our understanding of models as investigative instruments (Morgan & Morrison 1999), which enable scientists to learn about phenomena they know little or nothing about—and to do so efficiently.

2. Characterizing the problem of biomolecular structure determination

Let us begin by considering the mid-twentieth century problem of DNA structure and contemporary problems of biomolecular structure determination. In this section, I propose a way of understanding these problems and identify factors contributing to solving them efficiently.

In the late 1940s, DNA was known to consist of either two or three polynucleotide chains and four bases: the purines, adenine and guanine, and the pyrimidines, cytosine and thymine. What remained to be determined was how these component parts fit together. Were there two or three polynucleotide chains, and how were they connected to one another? What folded conformation did the molecule adopt? Searching for the structure of DNA was like attempting to assemble the pieces of a very complex puzzle. Those pieces could only fit together in particular ways; clues for assembling them came from experimental data such as X-ray diffraction photographs and theoretical considerations such as stereochemical principles.

Contemporary structural biologists, too, aim to determine structural models by taking empirical and theoretical information to be constraints on permissible structures. They have at their disposal a greater amount of information than Crick and Watson did. Notably, in addition to X-ray crystallography, solution nuclear magnetic resonance (NMR) spectroscopy and cryo-electromagnetic tomography can produce atomic-resolution models of biomolecular structures. Further, the Protein Data Bank (PDB), an online database that houses hundreds of thousands of previously determined protein and nucleotide structures, gives researchers easy access to information about related molecules (Berman et al., 2000). Contemporary structural biologists also tackle more complex problems. For example, the 52-MDa Nuclear Pore Complex (NPC) contains about 550 subunits of approximately 30 types (Alber et al., 2007). But the essential task remains the same: to determine a structural model with the help of theoretical and empirical information or, in other words, to find a model that accommodates available information sufficiently well (Rout & Sali, 2019; Sali, 2021).

Both the DNA and contemporary biomolecular structure determination cases can be characterized as proceeding via the successive elimination of portions of a space of possible structures through taking into consideration different pieces of empirical and theoretical information, each of which constrains the structure in particular ways.[1] For instance, if one had evidence that DNA contains two rather than three strands, one could eliminate all three-stranded structures from further consideration. If one then learned that these strands were connected by the bases, one could eliminate all structures that had the bases radiating outwards (Bolinska, 2018).

Of course, this is not a neat, linear process—far from it. Instead, different pieces of information had to be considered merely *possible* constraints on the structure: each was liable to mislead with respect to what it was taken to indicate about it. This is precisely what happened when Lawrence Bragg, John Kendrew, and Max Perutz attempted to solve the problem of polypeptide chain folding a few years before Crick and Watson were working on DNA. Bragg et al. (1950) were heavily influenced by an X-ray diffraction photograph of the protein keratin taken by William Astbury in 1932. This photograph was widely believed to indicate that the polypeptide chain would have a structure that repeated every 5.1 Å. In fact, however, the structure—which Linus Pauling determined a few years later—was found to have a subunit repeating every *5.4* Å. It turned out that the spot on the photograph seemingly resulting from a repeating subunit every 5.1 Å was instead caused by what became known as a 'coiled coil' higher-order structure (Judson, 1996). That is, there was an unexpected alternative interpretation available for what Astbury's photograph said about the structure of the folded polypeptide chain. In contemporary biomolecular structure determination, too, each piece of information is merely a possible constraint on structure: data are often sparse, noisy, or ambiguous, and therefore difficult to interpret with respect to what they indicate about the target system (Schneidman-Duhovny et al., 2014).

The process of structure determination, then, is an iterative one, proceeding by consideration of hypothetical structures suggested by combining different constraints: *if* we take these pieces of experimental data or theory as constraining the structure in such-and-such way, what follows? In particular, do we get a structure compatible with other information? If not, then we need to reconsider how constraints have been applied; perhaps we have made an error in interpretation or application somewhere along the line. Each iteration allows us to revise some of the assumptions we made in the previous one. The process continues until a model sufficiently compatible with available information has been found.[2]

An efficient strategy for determining biomolecular structures is one that, as much as possible, constrains the structure *in the right way* through the consideration of each piece of information. It reduces the likelihood of becoming misled by some piece of information, eliminating the correct structure from the possibility space (as Bragg, Kendrew, and Perutz did) and necessitating starting the process all over again to determine where one went wrong. Such a strategy minimizes how many iterations are required, on average, to get the right solution by reducing the likelihood of having to backtrack on any given occasion (Bolinska, 2018). Further, an efficient strategy directs one's attention toward parts of the possibility

[1] 'Information' here is understood broadly, as any bit of theory or empirical evidence that might be relevant to what the structure looks like. Information in this sense has two central properties: it is *extensive*, such that two independent datasets containing the same amount of information contain twice as much information as each alone; and it *reduces uncertainty* (Adriaans, 2020). Thus, the sense of information at hand is aligned with Shannon information (Shannon, 1948; Shannon & Weaver, 1949). See also Suárez and Bolinska (2021).

[2] Framing the process in this way leaves open the question about whether science is in the business of finding the truth. For a realist, a model most consistent with the available information is most likely to be true; for an anti-realist, consistency with information is a mark of empirical adequacy.

space in which the correct structure is likeliest to be found. Because the space of possible structures is so large, it could take a long time to search its entirety.[3] Without an efficient search strategy, one might end up searching indefinitely. This is particularly true the greater the size of the possibility space, and therefore even more important in contemporary structural biology.

3. Epistemic expression in biomolecular structure determination

With an understanding of biomolecular structure determination as the successive narrowing-down of a space of possible structures by considering information of varied reliability in place, let us turn to the role that epistemic expression plays in this process. I will argue that epistemic expressions share two common features: (i) they contain information in a way that enables the most reliable information to place the most stringent constraints on permissible structures (Section 3.1) and (ii) they facilitate the ready extraction of the information sought by investigators by biasing the search through the possibility space toward places in which the correct structure is likeliest to be located (Section 3.2).

3.1. Containing information

Recall that Bragg, Kendrew, and Perutz were misled by Astbury's X-ray diffraction photograph of keratin, proposing a solution that accorded with the common interpretation of that photograph as indicating a 5.1 Å repeat in the structure. But their structure erred in another way, too: it violated a stereochemical constraint. Due to a phenomenon known as resonance, the peptide bond, typically depicted in structural formulae as a single bond, in fact has partial double-bond character. Therefore, this bond is planar; rotation about it is prohibited. Yet the structure that Bragg, Kendrew, and Perutz selected for the folded polypeptide chain permitted rotation about the peptide bond (Olby, 1974).

Bragg, Kendrew, and Perutz's fundamental mistake was to rely too heavily on Astbury's photograph and not enough on information about bond lengths, bond angles, and stereochemical rules in directing their successive possibility space reduction. The former information was less reliable than the latter. X-ray diffraction photographs were amenable to several interpretations, which could be based on mistaken assumptions. On the other hand, precise values for bond lengths and angles had been determined and refined over years using a variety of experimental methods, and stereochemical principles held the status of highly confirmed theory (Bolinska, 2018). Thus, it was unlikely that the correct structure would be incompatible with these pieces of information; they were "constraints that the final answer *had in any case* to satisfy" (Crick, 1988, p. 60; emphasis added).

Crick and Watson's model-building served as an epistemic expression because the pieces from which they built their models were constructed precisely to scale, and thereby contained information about bond lengths, bond angles, and stereochemical rules. Any model that could be built from the pieces thus *automatically* took this information into account: it was impossible for it to get a bond length or angle wrong, or to violate a stereochemical rule, as Bragg, Kendrew, and Perutz's model of protein did. Moreover, the model omitted less reliable information—such as that contained in X-ray diffraction photographs—altogether, enabling hypothetical structures that were incompatible with that information to nonetheless be taken seriously as candidates for the correct solution. Taking seriously structures that did not accommodate less reliable information as candidate solutions decreases the likelihood of missteps like Bragg, Kendrew, and Perutz's. Indeed, Bragg, Kendrew, and Perutz considered a structure similar to Pauling's correct structure for the folded polypeptide chain (the alpha helix), but ruled it out because it was incompatible with the common interpretation of Astbury's photograph (Bragg et al., 1950). Had they accorded this incompatibility less importance, they might have realized that structure's other virtues.

By automatically taking information about bond lengths, bond angles, and stereochemical considerations into account, model-building put certain structures (those that could be physically constructed from the pieces) up for consideration and ruled others (those that couldn't) out. Because this information was reliable, model-building constrained which solutions were possible *in the right way*, making errors less likely. Model-building thus contributed to the efficiency of Crick and Watson's determination of the DNA structure, since backtracking to determine the source of an error can be computationally costly.

Whereas the epistemic expression in the case of DNA structure determination contained *only* the most reliable information, the more sophisticated machinery of integrative modeling permits information to constrain models *to varying degrees*. The aim of integrative modeling is to take all available information—from theory, experiment, and other models—into account to construct models of biomolecular structures that are sufficiently precise for answering biological questions. Model construction proceeds in three steps: defining the model *representation*, constructing a function for *scoring* alternative models, and *searching* a space of candidate models (Rout & Sali, 2019; Sali, 2021). Information can shape different steps, constraining models to varying degrees depending upon where and how it is used. Below, I describe what each step involves. I show that the most reliable information should be used to define the model representation and carry out searching, whereas less reliable information can be used for scoring.

The first step of integrative modeling is to define the model representation,[4] which specifies the mathematical variables whose values will be determined by modeling. For example, a common aim of modeling is to determine positions of individual atoms; in such a case, the model representation consists of x, y, and z coordinates for each atom in a structure. Alternatively, atoms and larger system components can be fixed with respect to each other into rigid bodies corresponding to previously determined structures.[5] Crick and Watson can be understood as having done just this in building models from pieces representing larger components of DNA, such as the bases adenine, guanine, cytosine, and thymine.[6] Their model pieces took for granted the atomic structures of these components, which had already been experimentally determined; the remaining question was how they were positioned relative to one another. Thus, the variables in Crick and Watson's model representation were the positions of these larger components, rather than of individual atoms.

The model representation effectively defines a space of in-principle possible models, with each model in the space specifying values for each of the model variables. In the DNA case, for example, the space of in-principle possible models included all possible arrangements of the model components with respect to one another. Defining the model representation places very stringent constraints on which models can be found: a model that isn't included in the space of in-principle possible models is thereby excluded from further consideration. Using unreliable

[3] If we allow the bond lengths and angles in candidate structures to adopt any real value, the search space is infinitely large. However, given limitations of the precision of our instrumentation, we can understand their values as being discrete rather than continuous (Bolinska, 2018).

[4] Note that 'model representation' is technical terminology, not to be confused with the more general notions of representation that philosophers use, such as the ones discussed in Section 4.

[5] A set of atoms can also be represented by a larger sphere, such as a bead corresponding to an amino acid residue or even a whole protein subunit. The model representation can specify the trajectory of a single system over time, or a system that exists in multiple states. For other examples, see Box 1 in Rout and Sali (2019).

[6] Indeed, Crick and Watson's model-building can be understood as an early instance of integrative modeling, in the sense that they aimed to take into account information from multiple theoretical and experimental sources in the determination of the DNA structure (Rout & Sali, 2019).

information to define the model representation therefore has serious potential to mislead: it can eliminate the correct model from consideration at this early stage.

I argued that, in the DNA case, the shapes of constituent parts were well known and highly confirmed, so it was prudent for these to constitute the system components of DNA, the positions of which were to be determined by modeling. We can now extend this point to contemporary integrative modeling: just as only the most reliable information was (rightly) contained in the model pieces for constructing a model of DNA, so too should only the most reliable information be used to define the model representation in integrative modeling. For example, in determining the structure of the Nuclear Pore Complex (NPC), researchers took for granted the structures of nucleoporins, key components of the NPC, in defining the model representation (Alber et al., 2007). The structures of nucleoporins had been previously experimentally determined, so using them in defining the model representation was likely to help narrow down the possibility space correctly, just as using the components in the DNA modeling case did for Crick and Watson. In contrast, relying on less reliable information to define the model representation would have the potential to mislead, in the same way that Bragg, Kendrew, and Perutz had been misled by the 5.1 Å spot in Astbury's photograph.[7]

Once a space of in-principle possible models has been defined by specifying the model representation, the next step of integrative modeling is to *score* each of these models with respect to how well they accommodate all of the input information. Most commonly, a least-squares scoring function is used, corresponding to a weighted sum of spatial restraints:

$$S = \mathbf{\Sigma}_i \, \omega_i \left(X_i \; - X_i^o \right)^2,$$

where the sum runs over all spatial restraints i, X_i is the value of a restrained spatial feature in a model, X_i^o is its measured value, and ω_i is the weight of the restraint. For example, a restraint (i) based on an NMR spectrum may compare the distance between two specific atoms in a model (X_i) with an experimental observation that this distance is less than 4.5 Å (X_i^o), weighted by our relative confidence in the measurement (ω_i). Each restraint thereby quantifies how much a computed property of a model deviates from what the input information specifies. By design, minimization of S minimizes the difference between the model and information about it (Rout & Sali, 2019; Sali, 2021).

In the case of DNA, any information that was less reliable than bond lengths and angles and stereochemical rules was omitted from model-building altogether. The scoring step in contemporary integrative modeling, however, gives us the resources to use information of intermediate reliability to constrain which structures are considered. The scoring function S sums over all spatial restraints, weighted according to our relative confidence in them. It thereby permits less reliable information to serve as a constraint on the structure to a lesser extent than information used to define the model representation. Whereas using information to define the model representation *excludes* any models incompatible with that information from further consideration, scoring permits information to constrain which structures are possible to a degree commensurate with its reliability. The weighting factor ω_i enables quantifying how important it is to accommodate different pieces of information, such that more weight can be placed on more reliable information. Contemporary integrative modeling therefore permits more degrees of freedom in how information of varying reliability can constrain which models are permissible than the DNA case does.[8]

The first condition on epistemic expression, then, is that the most reliable information place the most stringent constraints on which structures are permissible. Less reliable information may constrain structure if representational resources allow—as in the case of contemporary integrative modeling—but it must do so to a lesser degree, permitting consideration of structural models that deviate from it somewhat.

Let us turn now to the second condition on epistemic expression: that certain new information—namely, the information sought by investigators—be rendered readily available to them. What does it mean to make information readily available in the DNA and contemporary integrative modeling cases?

3.2. Extracting new information

In the previous section, I showed that the most reliable pieces of information can be used to define the model representation, where in the DNA case, this amounts to specifying the pieces from which models will be constructed and in contemporary integrative modeling, the variables whose values will be determined. I also showed that defining the model representation effectively specifies a space of in-principle possible models.

The next step of the modeling process is to search that space for the correct structure. Epistemic expressions expedite the extraction of information sought by investigators by biasing the search process toward regions of the space in which the correct structure is most likely to be found.

In the case of DNA, searching took place manually, via the construction of models from component pieces. The medium from which the pieces were made biased the search through the space by rendering certain models easier to construct than others. Those structures that were impossible to construct were ruled out immediately; they couldn't even be taken into explicit consideration. Other structures were possible to construct, but only with some difficulty. For instance, a model of DNA that had like-with-like base pairing would pinch in some places and bulge in others. This mechanical strain in the model corresponded to steric strain in the molecule (Charbonneau, 2013), enabling the physical medium of the model to bias Crick and Watson's search through the space of *prima facie* possible models toward places in which the correct structure was likeliest to be found.

Not all ways of representing information have this feature. For instance, a table listing bond lengths and angles would contain the same information as the physical model pieces did, but it would not offer the same searching bias that building models from precisely scaled pieces could. Information represented in a table would instead necessitate performing extensive numerical calculations. According to Pauling, who inspired Crick and Watson to adopt the model-building method, such calculations would be "so complex as to resist successful execution" (Judson, 1996, p. 63). In other words, failing to use an epistemic expression in this case might have precluded solving this problem (in practice, if not in principle).

In contemporary integrative modeling, searching is a partially automated process. Recall that the scoring function S quantifies how consistent a model is with all input information. The next step of integrative modeling, then, is to *search* the space of *prima facie* possible models for *acceptable* models, those that are sufficiently consistent with input information. In principle, a systematic enumeration generating every possible model one by one with sufficient granularity would be most thorough. However, enumeration is rarely computationally feasible, given the size of biomolecular structures and the precision required to enumerate them. So stochastic sampling methods, such as various Monte Carlo schemes, can be used instead. These methods aim to map the shape of the scoring function landscape[9] as a function of all model variables

[7] See also Bolinska and Sali (2023).

[8] See also Bolinska and Sali (2023).

[9] The notion of a scoring function landscape can be illustrated by a simplified example. A model representation that included only x and y coordinates for a single component of a modelled system would generate a space of in-principle possible models that included all possible permutations of x and y coordinates for that component. In such a case, this landscape would consist of x, y, and S variables.

without enumerating each model. They rely on heuristics that instead bias the search toward models that are more likely to be acceptable (Rout & Sali, 2019; Sali, 2021).

Rather than constraining the search physically, then, stochastic sampling methods rely on algorithms designed to preferentially search portions of the space of possible structures that are more likely to contain the correct structure. Further, they may also take into account information about the structure that wasn't used in the representation and scoring steps of modeling. For instance, in their determination of the position of a membrane protein in the NPC, researchers limited their search to the membrane, rather than searching the whole space of possible structures. The information that a particular protein was located in the membrane could be used to bias the search toward places in which the correct structure was likeliest to be found (Alber et al., 2007).

Limiting searching in these ways made new information readily available by isolating the most likely part of the possibility space in which the correct structure might lie. The alternative, searching by enumeration, might be so computationally expensive as not to be practically feasible, just as attempting to narrow down the DNA structure without the aid of a biasing mechanism such as model-building would be. By expediting the search through the possibility space, epistemic expression can render a problem solvable in practice that might otherwise be solvable only in principle.

3.3. Summary and key features

Determining biomolecular structures can be understood as proceeding by elimination of candidate structures based on different pieces of information, some of which are more reliable than others. This section showed how epistemic expressions can serve as direct routes to the identification of the correct structure from a vast space of possible structures. We considered the role of epistemic expression in two cases: Crick and Watson's determination of the DNA structure (via molecular model-building) and contemporary biomolecular structure determination (via integrative modeling). I argued that epistemic expressions contain information about the things they represent in a manner that ensures that the most reliable information places the tightest constraints on which structure is permissible. They further render new information readily extractible by biasing the search through the space of possible structures toward places in which the correct structure is likeliest to be found.

With this framework in place, let me highlight some central features of epistemic expression. First, epistemic expressions enhance research efficiency, but that doesn't mean they are infallible. The express post wouldn't be the express post if it consistently delivered letters at the same rate or slower than the standard post. This isn't to say that it always does so—it can get held up from time to time—but rather that, all else being equal, it is quicker. Its quick delivery is a product of procedures that are, on average, more efficient. Similarly, epistemic expressions are representations that tend to enhance research efficiency. Because research efficiency depends on numerous complex and interdependent factors, they might not always succeed in doing so. Nonetheless, epistemic expressions' tendency to enhance research efficiency is a product of their having the right features—namely, their enabling information to constrain structure to an extent that is warranted by its reliability and making the problem's solution salient by facilitating an effective search of the space of possible structures.

Second, what counts as an epistemic expression is relative to investigators' aims and cognitive capacities, as well as to how much and which information they have at their disposal. Here, we have been discussing the determination of biomolecular structures, but epistemic expression could be used for other epistemic aims such as prediction or understanding. Indeed, often biomolecular structures are determined with other goals in mind, such as predicting a drug's mechanism of action or understanding the molecular basis of heredity. Moreover, these goals can determine which information is most important and the resolution of representation required to meet them. For instance, understanding catalysis typically requires atomic-level resolution at the active site residues of an enzyme, whereas a coarser-grained representation might be sufficient for other purposes. Finally, epistemic expressions in cases in which investigators have little information available will by necessity contain less reliable information; that is the best we can do in cases of exploratory research. Because of this context-sensitivity, what qualifies as an epistemic expression must be considered on a case-by-case basis.

Third, relative to each investigative context, there are better and worse epistemic expressions; that is, epistemic expressions can expedite the efficiency of scientific research strategies to varying degrees. For example, while Pauling was sick and confined to bedrest, he decided to tackle the problem of polypeptide chain folding. He began by drawing out the polypeptide chain on a piece of paper (Fig. 1). Folding and refolding the paper along the peptide bonds, he searched through the space of possible structures, trying out different pitches for the helix (Judson, 1996). The material of the paper, together with the rough sketch of the polypeptide chain, contained information about bond lengths and angles and stereochemical rules. However, these bond lengths and angles were not drawn precisely to scale, and the material of the paper could not bias the search through the possibility space to the same extent as physical models could.

But, importantly, not anything goes. There are some forms of representation that simply do not increase efficiency or may even detract from it; they do not qualify as epistemic expressions. For instance, suppose that, rather than constructing models from precisely scaled pieces, Crick and Watson had instead taken blocks of Lego and randomly assigned each to a component of the molecule, trying to determine the structure by building a model from them. This model would not help them to find the structure more quickly because it would not contain the relevant information about bond lengths and angles.

We should therefore think about epistemic expression as a success term in the same way we do explanation. There are better and worse explanations, but when we ask questions about the nature of explanation, we are primarily interested in what makes an explanation a good one. We acknowledge that some explanations are better than others, and that some putative explanations are so bad that they are not explanatory at all—they do not in fact qualify as explanations. Similarly, here I am interested in what makes epistemic expressions effective. There can be epistemic expressions that are somewhat less effective, but still enhance research efficiency. But not all representations are epistemic expressions, nor are all cognitive activities in which representations are used instances of epistemic expression.

This point contrasts discussions about the nature of so-called scientific representation, which many authors do not regard to be a success term in this way. We consider some of these discussions in the next section.

4. Do we need an account of epistemic expression?

I have proposed an account of epistemic expression, a kind of representation that helps facilitate efficient scientific research strategies. But is such an account necessary? It might be thought that we already have answers to the questions this paper addresses. After all, a widely acknowledged aim of accounts of *scientific representation* is to explain how, by way of representing their target systems, models enable their users to learn about those systems (Bailer-Jones, 2003; Frigg, 2006; Frigg & Nguyen, 2020; Poznic, 2016; Weisberg, 2013). Some analyses of scientific representation make their commitment to this aim explicit, adopting alternative terms alongside or instead of "scientific representation." For instance, Mauricio Suárez (2004) takes scientific representation to be a form of "cognitive representation," where cognitive representations are those that can facilitate inference-making or *surrogative reasoning* (Swoyer, 1991) about their target systems. Others favor "epistemic representation," highlighting the knowledge-gathering role of this class of representation (Bolinska, 2013, 2016; Contessa, 2007; Frigg

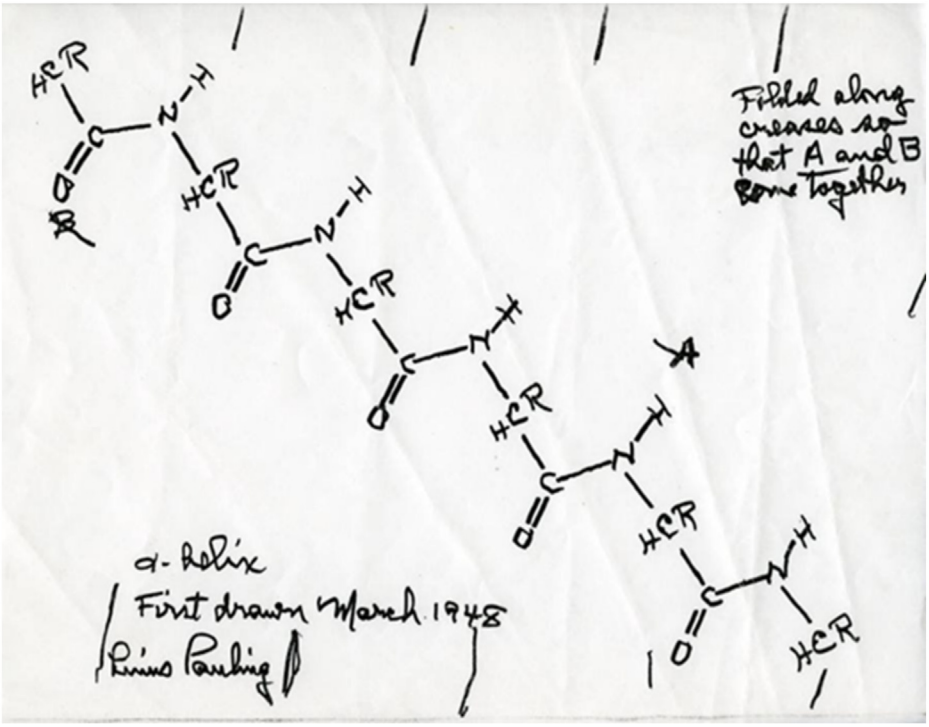

Fig. 1. Pauling's 1982 reproduction of his original (1948) drawing of the polypeptide chain. Ava Helen and Linus Pauling Papers, Special Collection & Archives Research Center, Oregon State University Libraries.

& Nguyen, 2020; Shech, 2015) or describe models as artifacts that function as *epistemic tools* (Knuuttila, 2005, 2011, 2017).

In this section, however, I argue that they do not in fact have these answers. To be clear, my aim is not to suggest that they *should* have them or to criticize them for failing to meet their explanatory goals, nor is it to suggest that they are fundamentally incompatible with epistemic expression.[10] Rather, it is to address the plausible suggestion, outlined above, that they might be readily extended to obviate the need for a distinct concept of epistemic expression. I show that these accounts, by themselves, cannot tell us how certain representations can help researchers solve problems more efficiently than they would have been able to without them.

To see why, I will highlight a tension that arises when we try to account for learning from models while also regarding models that misrepresent their targets as scientific representations (Section 4.1). I will show that two plausible suggestions for resolving this tension—via a pragmatic understanding of scientific representation (Section 4.2) or conceiving of models as artifacts (Section 4.3)—significantly weaken the sense of learning from models in question. Understanding models as epistemic expressions can help us to better understand the stronger sense of learning that is often imperative in science.

4.1. The tension between learning from models and the possibility of misrepresentation

Representations do not always succeed in accurately portraying their target systems, and a widespread view holds that this should not preclude them from qualifying as scientific representations. Suárez writes that "[o]n discovering particular inaccuracies in [a particular] representation we are very rarely inclined to withdraw the claim that it is *a* representation" (2003, 226). Permitting misrepresentation is commonly thought to be a condition of conceptual adequacy. For instance, according to Andreas Bartels, "the very concept of representation presupposes the possibility of a distinction between the case in which some X *misrepresents* some Y and the case in which X does *not* represent Y at all" (2006, 13). On such an understanding, misrepresentation is a species of representation; accordingly, an adequate account of scientific representation must encompass cases of successful representation and misrepresentation alike.[11] Again, note the contrast with much of the literature on explanation, which conceives of explanation largely as a success term.

A tension arises. On its face, the aim of devising an account of scientific representation that includes misrepresentation as a species seems incompatible with the aim of explaining how learning from models is possible. For if a model is a poor enough misrepresentation of its target, in what sense can we learn about the target by consulting the model?

4.2. Resolving the tension (i): pragmatic accounts of scientific representation

One way of resolving the tension is to argue that models enable the drawing of inferences about their targets, but that these inferences need not be true. This line of thinking is endorsed by proponents of *pragmatic* accounts of scientific representation (e.g., Contessa, 2007; Frigg & Nguyen, 2016, 2018, 2020; Giere, 2004, 2010; Suárez, 2004; Van Fraassen, 2008), those that emphasize the role of users and their purposes in the activity of representation, rather than narrowly focusing on the relationship between a model and its target.[12] It originates in R. I. G. Hughes's (1997) denotation-demonstration-interpretation (DDI) account of scientific representation, according to which we begin by *denoting* elements of the target with elements of the model; then we *demonstrate* that a particular conclusion is true by reasoning within the model; finally,

[10] Indeed, epistemic expressions are a subset of scientific representations that satisfy further conditions.

[11] See also Frigg (2006), Frigg & Ngyuen (2020), and Shech (2015).

[12] Views in this latter family have been variously referred to as *informational* (Chakravartty, 2010), *two-place* (Knuuttila, 2011), or *substantive reductive* (Suárez, 2010) accounts of representation. See, for instance, Van Fraassen (1980, 1989), French (2003), Da Costa and French (2003), and Bartels (2006).

we *interpret* our findings with respect to the target, inferring that the corresponding conclusion is true of it, too.

Hughes's DDI account has been extended and refined, for instance in Roman Frigg and James Nguyen's (2016, 2018, 2020) DEKI account. DEKI "places no restrictions on the choice of the vehicle" of representation (Frigg & Nguyen, 2016, p. 235) but holds that, in addition to *denoting* its target, a model must also *exemplify* certain features and come with a *key* enabling the *imputation* of at least one of these features to the target. A feature is exemplified when it is instantiated and highlighted—that is, selected as relevant and made epistemically accessible—in a model. Learning about the target system is made possible by the key:

> "[I]f one's preferred model is [to] be used to provide information about an actual system in the world, then one has to be explicit about how features of the model are supposed to correspond to features one conjectures the target to have. Ignoring this connection amounts to investigating the model without investigating any actual system in the world" (Frigg & Nguyen, 2020, 182).

In other words, the key is what connects the model to some system in the world, rendering it a model *of* that system. Yet keys can be "highly conventional" and need not rely on any similarity between vehicle and target (ibid.). And representation need not be accurate; indeed, the system "need not possess *any* of the features that are ascribed to it" by the model (ibid., 178; my emphasis).

The fact that pragmatic accounts of scientific representation are committed to permitting misrepresentation is precisely what makes them unsuitable for explaining epistemic expression. Consider again the case in which Crick and Watson build models from Lego blocks. This hypothetical model-building activity would meet the conditions of pragmatic accounts, even sophisticated ones like DEKI. Crick and Watson could use the Lego blocks to denote parts of the DNA molecule, and the model they could construct would denote DNA. The blocks and the models constructed from them could also exemplify certain features, such as being rigid, having a particular shape, and being connected to one another in certain ways. These features are instantiated in the pieces and highlighted in the resultant model; by construction, they can be selected as relevant features; and they are epistemically accessible. The model can come with a key associating the spatial relations among the pieces with spatial relations among the components of DNA, and these features can be imputed to DNA.

On the DEKI account, then, the Lego block model of DNA would qualify as a scientific representation of DNA. But it would not tell Crick and Watson much about the structure *of DNA*, and therefore would not enhance the efficiency of learning about its structure. For, given that the pieces do not resemble the components of the molecules in the relevant respects, the models constructed from them would not tell Crick and Watson anything about its structure. Instead, they would tell them about a *hypothetical* structure with components that have the shapes the model ascribes to them.[13] Contra DEKI, the existence of a key enabling the imputation of some features of the model to features of the target is not enough for the model to be a model of the structure.

This point highlights a widespread confusion. Following Nelson Goodman (1968), many authors insist that resemblance is not necessary for aboutness or reference, which can be established by stipulation (e.g., Callender & Cohen, 2006). This is a core component of the intuition that misrepresentation is a species of representation: representation is reducible to reference; anything can refer to anything else; therefore, even severely misrepresentational models can represent.

There is a sense of reference on which this is correct. We can indeed take anything to *denote* or stand in for anything else simply by stipulating that it does so; that is, anything can serve as what Liu (2015) calls a *symbolic* vehicle of representation of a given target. This is how conventional signs and symbols take on their meaning. But in modeling the structure of DNA, we do not want our model to be about that structure merely in this weak sense of just pointing to it. Our goal isn't to use the model in place of "structure" when talking about the structure, or as a symbol directing our attention to it. Rather, we want it to be about the structure in a stronger sense: we want it to be an *epistemic* vehicle of representation (Liu, 2015), enabling us to learn what the structure is like from it. And the Lego block model does not enable us to do so. It is simply not about the structure in this stronger sense.

An analogy can illustrate the point. We could certainly stipulate that a portrait of Rafael Nadal refers to Roger Federer, and then use the Nadal portrait to talk about Federer in certain contexts. For instance, we could use it as a token on a virtual tennis court to demonstrate how Federer moved across the court in his last match. The Nadal portrait, as a result of our stipulative act, comes to refer to Federer in this context. But there is a distinct and common understanding of aboutness in which the Nadal portrait cannot be about Federer because the Nadal portrait doesn't have the right representational content. Although we can use it as a placeholder to refer to Federer in a context like the one described here, it's still a portrait *of Nadal*.[14] We cannot use it if we want to learn what Federer looks like. There are different senses of aboutness or reference: a weaker sense, in which aboutness can be established by fiat; and a stronger sense, in which some degree of resemblance is required. When we want to learn about what something is like, rather than just pointing to it, the weaker sense of reference as denotation is not enough.

This is not to say that the stronger sense of aboutness requires complete resemblance; it instead permits misrepresentation to a certain degree. For instance, we could still learn something about what Federer looks like from a caricature of him that exaggerated some of his features.[15] And of course there could be disagreement about what degree of misrepresentation is bad enough to preclude learning about his appearance. But the point is that mere stipulation, whereby a bona fide portrait of Nadal is taken to represent Federer, will not do. By requiring a degree of resemblance to their targets, epistemic expressions enable learning about them in a stronger sense than pragmatic accounts of scientific representation can accommodate.

Further, epistemic expressions do more than just enabling surrogative reasoning. We can reason about a system consisting of two ships moving along the surface of the sea by denoting each of the ships by a pen, and having the pens move along a piece of paper. We can then infer that the actual ships have similar trajectories on the surface of the sea as do the pens on the piece of paper. According to Suárez (2004), the fact that the pens-on-paper system refers to the ships on the sea, and that we can make inferences from the former to the latter, makes the pens-on-paper system a scientific representation of the ships on the sea. Reference and facilitating surrogative reasoning are what make the pens-on-paper system a scientific representation of the ships-on-sea system.

The case of epistemic expression is more complex. We do not simply denote certain elements of DNA structure with elements of the model, then build the model, and finally infer that DNA has a corresponding structure. Rather than merely *denoting* components of DNA, the model pieces *contain reliable information* about them. They do so by virtue of being constructed precisely to scale; the information they contain is thus sufficiently accurate for the purpose at hand. Similarly, in contemporary integrative modeling, information can be used either to define the model representation or for scoring alternative models with respect to how well they accommodate input information; it matters which information is

[13] This also isn't to deny that there are scenarios in which thinking about a hypothetical system can lead us to learn about an actual one, such as the modeling of perpetual motion machines or populations with three sexes to show why our world doesn't include them (Weisberg, 2007).

[14] As Elay Shech (2015) points out, on most accounts of linguistic and pictorial representation, reference is derived from the representational content.

[15] Indeed, misrepresentation can have a useful epistemic function, since it can serve to highlight important features of the target. This can be true both of caricatures and of idealized models.

used in which step. What's more, in both cases, information is presented in a form that biases the reduction of the space of possible solutions toward the correct solution. Surrogative reasoning is certainly involved, but it alone does not suffice for epistemic expression.

Thus, although pragmatic accounts of scientific representation seem at first glance to be able to answer questions about how certain representations can enhance the efficiency of scientific research strategies, they are ultimately unsuited to this purpose. Epistemic expressions are not mere surrogates for reasoning. Their connection to the targets they represent is not entirely conventional: it is vital that they contain information that reliably constrains the space of possible solutions to a research problem, and that they enable the ready emergence of that solution by biasing the search through this space.

4.3. Resolving the tension (ii): models as artifacts

An alternative approach to resolving the tension between learning from models and taking misrepresentation to be a species of scientific representation, championed by Tarja Knuuttila, is to understand models as *artifacts*, "external tools for thinking, the *construction* and *manipulation* of which are crucial to their epistemic functioning" (Knuuttila, 2011, p. 263).[16] According to Knuuttila (2011, 2017), models are epistemic tools that constrain and afford solutions to a problem by virtue of being concretely manipulable; they act as scaffolds for cognition by making important features of the system salient to their users. Although it builds on Morrison and Morgan's (1999) conception of models as investigative tools, Knuuttila's understanding parts ways with it in its rejection of the idea that models' representational role is necessary for explaining how we learn from them. Just as a hammer can be used to drive a nail into a wall without representing anything, so too can we account for the epistemic function of models "quite apart from any determinate representational ties to specific real-world target systems they might have" (Knuuttila, 2011, p. 267). This approach, then, takes a step further than pragmatic approaches that adopt a deflationary or minimalist attitude toward representation. Thinking about models as artifacts thus "loosens them from any preestablished, fixed and well-defined representational relationships" (Knuuttila, 2005, p. 1261).

Knuuttila's models-as-artifacts view, with its emphasis on the construction and manipulation of models, comes closest to being able to supply an understanding of epistemic expression. However, contra Knuuttila, I urge that certain determinate representational ties to target systems *are* required for epistemic expression. As I showed in Section 3.1, the success of Crick and Watson's model-building depended crucially on model pieces being constructed precisely to scale—that is, their accurately representing components of the molecule. Similarly, the success of contemporary integrative structure modeling depends upon information correctly constraining the possibility space in the representation and scoring steps. But the Lego model-building case, which is not an instance of epistemic expression, *could* qualify as an artifact in Knuuttila's sense: the Lego pieces could act as tools for thinking that constrain and afford solutions to the problem, being concretely manipulable and making certain features salient. Thus, being an artifact is not sufficient for being an epistemic expression.

Knuuttila holds that models "are valued often […] for what they produce [rather] than for being truthful representations of their (supposed) natural target systems," since "usually we do not know enough about those systems, which is exactly the point of modeling" (2005, 1268). What I have shown is that what models produce is in fact *dependent on* their being truthful representations of certain aspects of their target systems. Knuuttila is right that the point of modeling is often to learn about systems we don't understand well. But the success of modeling depends on our ability to correctly embody the information that we *do* have—even when we don't have much. Furthermore, as I have shown, given that we have varying degrees of confidence in the information we have, it is imperative that we choose the most reliable information to guide our efforts.

Thus, Knuuttila's models-as-artifacts approach, too, does not have the resources to explain how certain representations can enhance research efficiency. It cannot stand in for an account of epistemic expression.

5. Conclusion: models as investigative instruments reexamined

Discussions of the nature of scientific representation often overlook an important distinction: between contexts in which representations serve the purpose of communicating what's already known about a target and those in which they facilitate investigation of a poorly understood target (Bolinska, 2016). But this distinction, I maintain, is significant because it highlights different functions of models and therefore different ways of thinking about what is important about them.

When we already understand a system fairly well, we can worry about how best to convey what we know about it to others. But when we don't yet know very much, the challenge is wholly different: how do we *re*present something when we don't know how it is presented in the first place? We are faced with several layers of uncertainty—in what we do and don't know, and in whether and how to capture different pieces of information in our representations. Our task is to leverage what we *do* know to learn more. 'Leverage,' with its root 'lever', is an apt term here: how can we best use what we are most confident that we know well to increase most efficiently how much more we can learn about a system?

To answer these questions, I introduced the notion of epistemic expression, where an epistemic expression helps its users to solve a scientific problem most expediently. Contra pragmatic accounts of representation and the models-as-artifacts view, I argued that sufficiently accurate representation is essential for epistemic expression. What is more, *which* information a representation contains and *how* it does so is crucial: epistemic expressions must contain information in a way that enables more reliable information to constrain the possibility space to a greater extent. And they must present that information to their users in a way that makes the solution to their research problem readily extractible by biasing the search toward places in which it is likeliest to be found. Whereas Crick and Watson built physical molecular models to expedite the determination of the DNA structure, integrative modeling in contemporary structural biology is implemented in computer software. Despite the difference in medium and method, molecular model-building and integrative modeling both served as epistemic expressions in the determination of biomolecular structure.

This account of epistemic expression therefore extends Morrison and Morgan's (1999) conception of models as investigative instruments by saying more about how models' independence from and dependence on the things they represent is important for learning from them. Partial dependence matters because without it, it's not clear that we are learning *about* the system in question in the first place. Although models are partially independent from the systems they are about, they "must also connect in some way with the theory or the data from the world," since "otherwise we can say nothing about those domains" (Morrison & Morgan, 1999, p. 17). I've shown that the connection is a product of their containing sufficiently accurate information about reliably known constraints. Contra Knuuttila, certain preestablished, fixed, and well-defined representational relationships *are* required for successful epistemic expression. Partial independence, too, is important because it is what enables users to access information that would otherwise be inaccessible to them. In particular, epistemic expressions' form, be it the physical medium of molecular model pieces or the steps of integrative modeling implemented in software, enables new information—the solution to the research problem—to be extracted from the information at hand. Especially with the rise of big data in areas like biology (Leonelli, 2016), such partial independence is especially crucial. Together, these features explain how "models can have a life of their own" (Morrison & Morgan, 1999, p. 18) in the generation of scientific knowledge.

[16] See also Boon and Knuuttila (2009), Knuuttila (2005, 2011, 2017) and Knuuttila and Voutilainen (2003).

Acknowledgements

I am grateful to Andrej Sali for extensive discussions that have helped me to better understand integrative modeling. I presented parts of this paper at the 2021 meeting of the Canadian Society for the History and Philosophy of Science and the University of South Carolina Queen Street Symposium; I thank my audiences for their helpful feedback. For comments on drafts of this paper, I am grateful to Riana Betzler. Finally, I thank two anonymous referees from this journal for helping me to improve the paper and broaden its scope.

References

Adriaans, P. (2020). Information. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. URL https://plato.stanford.edu/archives/fall2020/entries/information.

Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Rout, M. P., & Sali, A. (2007). Determining the architectures of macromolecular assemblies. *Nature, 450*, 683–694.

Bailer-Jones, D. M. (2003). When scientific models represent. *International Studies in the Philosophy of Science, 17*(1), 59–74.

Bartels, A. (2006). Defending the structural conception of scientific representation. *Theoria, 55*, 7–19.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein Data Bank. *Nucleic Acids Research, 28*(1), 235–242.

Bolinska, A. (2013). Epistemic representation, informativeness and the aim of faithful representation. *Synthese, 190*, 219–234.

Bolinska, A. (2016). Successful visual epistemic representation. *Studies In History and Philosophy of Science Part A, 56*, 153–160.

Bolinska, A. (2018). Synthetic versus analytic approaches to protein and DNA structure determination. *Biology and Philosophy, 33*(3–4), 26.

Bolinska, A., & Sali, A. (2023). *Heuristics for integrative structure modeling: How to use information effectively*. Department of Philosophy, University of South Carolina [Unpublished manuscript].

Boon, M., & Knuuttila, T. (2009). Models as epistemic tools in engineering sciences: A pragmatic approach. In A. Meijers (Ed.), *Handbook of the philosophy of technological sciences: 9. Philosophy of technology and engineering sciences* (pp. 687–719). Amsterdam: Elsevier Science. No. IX).

Bragg, S. W. L., Kendrew, J. C., & Perutz, M. F. (1950). *Polypeptide chain configuration in crystalline proteins. 203A*. Proceedings of the Royal Society.

Callender, C., & Cohen, J. (2006). There is no special problem about scientific representation. *Theoria, 55*, 67–85.

Chakravartty, A. (2010). Informational versus functional theories of scientific representation. *Synthese, 172*, 197–213.

Charbonneau, M. (2013). The cognitive life of mechanical molecular models. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44*(4), 585–594.

Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science, 74*(1), 48–68.

Crick, F. H. C. (1988). *What mad pursuit: A personal view of scientific discovery*. New York: Basic Books.

Da Costa, N. C., & French, S. (2003). *Science and partial truth: A unitary approach to models and scientific reasoning*. Oxford: Oxford University Press.

French, S. (2003). A model-theoretic account of representation (or, I don't know much representation about art. but I know it involves isomorphism). *Philosophy of Science, 70*, 1472–1483.

Frigg, R. (2006). Scientific representation and the semantic view of theories. *Theoria, 55*, 49–65.

Frigg, R., & Nguyen, J. (2016). The fiction view of models reloaded. *The Monist, 99*(3), 225–242.

Frigg, R., & Nguyen, J. (2018). The turn of the valve: Representing with material models. *European Journal for Philosophy of Science, 8*(2), 205–224.

Frigg, R., & Nguyen, J. (2020). *Modelling nature: An opinionated introduction to scientific representation*. Springer.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*(5), 742–752.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese, 172*(2), 269–281.

Goodman, N. (1968). *Languages of art: An approach to a theory of symbols*. Indianapolis: The Bobbs-Merrill Company.

Heesen, R. (2018). Why the reward structure of science makes reproducibility problems inevitable. *The Journal of Philosophy, 115*(12), 661–674.

Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science, 64*(4), 325–336.

Judson, H. F. (1996). *The eighth day of creation: Makers of the revolution in biology*. Plainview: New York: Cold Spring Laboratory Press.

Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of Science, 72*, 1260–1271.

Knuuttila, T. (2011). Modelling and representing: An artifactual approach to model-based representation. *Studies In History and Philosophy of Science Part A, 42*(2), 262–271.

Knuuttila, T. (2017). Imagination extended and embedded: Artifactual versus fictional accounts of models. *Synthese, 198*(Suppl 21), S5077–S5097.

Knuuttila, T., & Voutilainen, A. (2003). A parser as an epistemic artifact: A material view on models. *Philosophy of Science, 70*, 1484–1495.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. Chicago: University of Chicago Press.

Liu, C. (2015). Re-Inflating the conception of scientific representation. *International Studies in the Philosophy of Science, 29*(1), 41–59.

Maddox, B. (2002). *Rosalind franklin: The dark lady of DNA* (1st ed.). New York: Harper Collins.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan, & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science*. Cambridge: Cambridge University Press.

Olby, R. (1974). *The path to the double helix: The discovery of DNA*. London: MacMillan.

Poznic, M. (2016). Representation and similarity: Suárez on necessary and sufficient conditions of scientific representation. *Journal for General Philosophy of Science, 47*(2), 331–347.

Rout, M. P., & Sali, A. (2019). Principles for integrative structural biology Studies. *Cell, 177*(6), 1384–1403.

Sali, A. (2021). From integrative structural biology to cell biology. *Journal of Biological Chemistry, 296*, Article 100743.

Schneidman-Duhovny, D., Pellarin, R., & Sali, A. (2014). Uncertainty in integrative structural modeling. *Current Opinion in Structural Biology, 28*, 96–104.

Shannon, C. E. (1948). A mathematical theory of communication, 379–423 & 27(4) *Bell System Technical Journal, 27*(3), 623–656.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Shech, E. (2015). Scientific misrepresentation and guides to ontology: The need for representational code and contents. *Synthese, 192*, 3463–3485.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science, 71*(5), 767–779.

Suárez, M. (2010). Scientific representation. *Philosophy Compass, 5*(1), 91–101.

Suárez, M., & Bolinska, A. (2021). Informative models: Idealization and abstraction. In A. Cassini, & J. Redmond (Eds.), *Logic, epistemology, and the unity of science: 50. Models and idealizations in science*. Cham: Springer.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese, 87*(3), 449–508.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.

Van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Clarendon.

Van Fraassen, B. C. (2008). *Scientific representation*. Oxford: Oxford University Press.

Weisberg, M. (2007). Who is a modeler? *The British Journal for the Philosophy of Science, 58*(2), 207–233.

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