4 Is Captain Kirk a natural blonde?

Do X-ray crystallographers dream of electron clouds? Comparing model-based inferences in science with fiction

Ann-Sophie Barwich

1 Introduction: science and fiction

Scientific models and fiction have one noticeable feature in common. Their representational relation to the physical world is ambiguous. It is often not obvious whether certain elements in a model represent something in the world or are an artifact of a model's internal structure. Fiction, too, can mimic our world to varying degrees, as fictional worlds sometimes contain historical characters or events, such as Henry VIII orr the Stonewall Riots.

Correction: "or" instead of "orr"

When we use scientific models, however, expectations of how our inferences address the world differ from interpretations of fiction. We consider scientific models to be representations that are true about something in the world. By contrast, we regard fictions as being important records of human culture, but not as true of anything in particular. Wherein is this difference grounded, and how is it justified?

Delete: "q" between "this" and "question"

The increasing dependency of scientific research on mediated forms of observation and depiction makes this q question central to philosophical interest about the characteristics of scientific representation. Laboratory conditions are hardly representative of many natural phenomena that we aim to investigate through them. That is true no matter whether we talk about the development and use of model organisms (Ankeny and Leonelli, 2011), set-ups in sensory measurement (Barwich and Chang, 2015), or studies of protein synthesis (Rheinberger, 1997). From this perspective, a large part of the philosophical debate has engaged with the deep dependency of scientific inquiry on indirect modeling practices. If modern science builds on a necessarily mediated approach, how must we understand its claim of giving us access to (the fundamentals of) reality?

In reply to this question, some philosophers have started comparing scientific representations with fiction to understand their potentially common basis as *cognitive strategies* (Suárez, 2009). Consider the case of idealizations or abstractions. These strategies provide us with scenarios of events or the behavior of entities in particular circumstances. In many cases, such as ideal gases or frictionless planes, these scenarios are not realized or realizable in the actual world. Is there something that distinguishes these scenarios from fictionalizations?

My focus in this chapter is on the special epistemic role that we assign to scientific representations as giving us indirect access to the reality of nature. A significant number of philosophical arguments have been directed at the dyadic relation between a model and its target system. Central to these arguments is an understanding of representations that addresses their structural features (e.g., isomorphism) or their conventional use as "as if scenarios." Here, I pursue an alternative tactic. Looking closely at model-building strategies, I focus on the interpretative strategies that deal with the representational limits of models. The chief question is, how do we interpret ambiguous elements in models? Moreover, how do we determine the validity of inference about information that is not explicit in a model? I suggest that the answer lies in the particular strategies that link a model to other methods in an experimental context.

In what follows, the chapter begins in Section 2 with the problem of representation in the contemporary philosophy of science. After introducing the reasons that prompted a comparison of scientific models with fiction, I argue that the problem of ambiguous inference emerges from two essential features of representations, namely their hybridity and incompleteness. To distinguish between fictional and non-fictional elements in scientific models, my proposal is to look at the integrative strategies that link a particular model to other methods in an ongoing research context. To exemplify this idea, I examine protein-modeling through X-ray crystallography as a pivotal method in biochemistry. As many readers from the humanities and the arts may not be familiar with this method, section 3 introduces the procedure in greater detail. My reason for this is to allow the reader to judge whether she considers my concluding analysis of the underlying fictionalization strategies in section 4 as adequate.

2 Context and argument: the problem of representation

Substitute: "depicts" instead of "represents"

What are the origins of the philosophical debate about whether a scientific representation represents reality accurately? Besides, what epistemological concerns suggest a comparison of scientific models with fiction? These questions have an inevitable historical dimension, so three main factors must be pointed out briefly. First, there is the legacy of radical theory changes, especially in nineteenth-century chemistry and twentieth-century physics. These changes shattered the epistemic confidence in our scientific methods and models. How can we be sure that our current theories and representations are more truthful than those left to rot in the graveyard of scientific history? Will our current concepts and models fail us, too? This issue is known as the pessimistic meta-induction in the philosophy of science, and if history is any indicator, a cautious attitude towards proper candidates for truth is advisable.

Second, the rise of studies that recognized science as an essentially social and historical endeavor further substantiated this caution. Social studies of science, especially over the second half of the twentieth century, demonstrated the contingency of factors that shape scientific advancement (Longino, 1990; Barnes, Bloor, and Henry, 1996). Third, when contemporary philosophy picked up

Delete: "philosophical"

on these developments, it turned away from broad generalizations about the nature of science. Instead, philosophical scholars directed their attention to the cognitive and experimental practices that underlie scientific research, for instance, in the construction of scientific concepts and representations such as models (Morgan and Morrison, 1999; Nersessian, 2010).

One line of investigation emerging from these advancements is the recent interest in the problem of representation (Knuuttila, 2005). As artifacts of human activity, how can scientific representations provide us with access to the world? How do we use these representations to gain knowledge about real things? Moreover, how must we understand cases of representational failure? On what basis are our model-based inferences justified? These questions have occupied many philosophers over the past decade. Their arguments revealed that scientific representations, as cognitive tools, often rely on fictionalization strategies or features that are shared by forms of fiction.

The meaning of fiction in this context is that of mimicry and distortion (Frigg, 2010a). Fiction as mimicry refers to representations that are designed to resemble real phenomena without truly referring to them, or without claiming to be a proper or truthful representation of these phenomena. In comparison, fictions as distortions describe alterations that present an (intentionally or unintentionally) alienated or converted image of a phenomenon. A differentiation of these two meanings, mimicry and distortion, is not necessarily clear-cut. Satire is a wonderful example of the thin conventional – and legal – line that holds between the mimicry of something (or somebody) in parallel with the explicit distortion of its features. Overall, this understanding of fiction centers on two fundamental philosophical topics: the reference of representations (or its suspension) on the one hand, and the truthfulness or accuracy of representations on the other.

Corresponding to this idea of fictionality, some philosophers of science encountered similar issues in the analysis of representations in science (for a collection of essays, see Suárez, 2009). Almost all scientific models build on various forms of distortions, abstractions, idealizations, imaginative scenarios, metaphorical comparisons, analogies, and so on (Hesse, 1966; Holyoak and Thagard; 1996; Van Fraassen, 2010). The influential "billiard ball" model of the atom by Dalton, the rise of the ball-and-stick model of chemical substances by August Wilhelm von Hofmann, or the Homo Economicus in economic theories are just some of the many prominent examples. Furthermore, many scientific entities were believed to be true initially, but turned out not to exist. Consider the abandoned scientific concepts of the ether, pneuma, phlogiston, or the idea of a *spiritus rector*. Nevertheless, these entities yielded experimental results and temporary insights into some aspects of the world at first, as it has been argued convincingly for the case of phlogiston (Chang, 2012).

Delete: "at first"

However, the comparison of science and fiction is not immediately intuitive. It does not come as a great surprise to scientists that we should not take scientific representations as literal depictions of the world. Neither does it sound astonishing that science rests on a graveyard of theoretical entities

and proceeds through self-correction by its very nature. Science continuously changes and expands the scope of our knowledge by pointing at what we do not know. From this view, it is an infinite business (Deutsch, 2011). To be sure, these constitute central philosophical insights into the nature of science that have been gained over the past century. So, what does the comparison of science with fiction tell us from a contemporary perspective?

Insert: "yet" between "we" and "do"

It stands to reason that science and fiction are not the same things. We do ascribe scientific representations a different epistemic status. They tell us something about what is real. However, the appropriate grounds for this claim are not always clear. In response, there are several angles from which we can approach the special status of scientific representations.

One way is to compare the structural features of scientific and fictional representations. What constitutes the resemblance or the similarity between a representation and its target system? Also, what are the functions of distortions in this context? Some philosophical arguments defined representation as a straightforward mapping connection that relates a model structure to a physical structure. An example is isomorphism: "A Model M is a structure; and M represents a target system T if T is structurally isomorphic to M" (Frigg, 2002). By contrast, fiction does not seem to rest on such a correspondence. On this account, structure as a representational criterion is based on a dyadic comparison between models and target systems.

Meanwhile, several arguments have shown that such a comparison does not lead to a clear demarcation of non-fictional from fictional representations. Criteria that seek to define the capacity of models to represent a target system through purely structural criteria must fail eventually (Goodman, 1969; Frigg, 2002). My view on this issue is that criteria of "similarity" and "structure" (whatever these may be) are indeed insufficient. Though, I cannot help thinking that the use of structure in these arguments presents a straw man.

Noticeably, this approach runs into trouble if we consider models as conglomerates and as consisting of different model ingredients (Boumans, 1999). From this perspective, the representational function of structures is contingent on the context of model-building. This point of view led to an emphasis on the epistemic context in which scientific representations are used (Knuuttila and Voutilainen, 2003). Here, the representational capacity of models and other scientific representations, such as diagrams or algorithms, is defined primarily by their epistemic function instead of their structural correspondence. Representations are understood to act as vehicles for reasoning and for making indirect inferences (Suárez, 2004). This means that representations are part of an imaginative act of "make believe." They present us with "as if scenarios" that give us "fictional" as in model-dependent truths (Walton, 1990; Frigg, 2010b; Toon, 2012). This idea strongly relies on a conventionalist understanding of an institutional or collective agreement about the use of representations (Searle, 2010).

Insert: "Thus, " before "representations"

However, such reference to the use and conventions of an epistemic context seems to beg the question. It presupposes that we already know how to use

(parts of) a representation adequately (i.e., as denoting or non-denoting). In fact, this account even dodges the real issue of a comparison between fiction and non-fiction as long as it avoids an answer to the question about the particular character of scientific representations. In essence, on what grounds is something referred to as a non-fictional representation? How can we find out whether an inference is only a "fictional truth," as in an essentially model-dependent truth, or whether it accounts for something model-independent but real?

In reply to this question, I propose an alternative tactic. Representations are neither judged by structural features in a dyadic comparison nor seen as make-believe scenarios. Rather, my focus is on the particular strategies that we employ in the interpretation of representations, and how these strategies help us to distinguish between what may be a fictional (model-dependent) and non-fictional (representational) model-component. My argument consists of two claims about the ontology of representations, fictional as well as non-fictional.[1] These are as following: First, most representations, fictional and non-fictional, are hybrids. Second, every individual representation is incomplete.

Beginning with the first claim, hybridity means that most representations contain denoting as well as non-denoting elements (for a more detailed argument see Barwich, 2013). For example, Bulgakov's fictional novel *The Master and Margarita* tells us a story about the devil having a ball in Moscow of the 1930s. Theological disputes aside, the devil is not what most people would consider a real person today. Nonetheless, there was such a place as Moscow of the 1930s. For a complementary example, consider the ball-and-stick models of chemical substances as a scientific representation. These models account for the basic spatial organization of a molecule, but in a highly idealized sense. In light of such mixed characteristics, it is often not intuitive what individual elements in a representation denote and what do not. Moreover, do these elements denote independently of the overall epistemic status of the representation in which they are contained (i.e., does it matter whether these elements are part of a novel or a scientific model)? For example, does the fictional Napoleon in *War and Peace* refer to the real historical figure of Napoleon? More so, what if we encounter a fictional Napoleon that has little in common with our knowledge of the historical figure? What about the pig Napoleon in George Orwell's *Animal Farm*? Besides, are non-existent entities that are part of working scientific models (such as silogens) fictions?[2] Furthermore, can fictional elements be accidentally true? Consider the possibility that we find a real person that matches every description of a fictional character's biography – without the author's (and perhaps even that individual's) knowledge. Is the fictional character suddenly a true description of this real person? There have been different answers to this problem, so it seems that the response depends on one's personal philosophical predilections (Ryle, 1933; Danneberg, 2006). Any position on this issue appears to presuppose a specific understanding and definition of what the notions of fictional and non-fictional entail. A clear demarcation between fiction and scientific representations as based on their denoting and non-denoting elements is inadequate.

Insert: "the" between "in" and "Moscow"

Insert: "the" between "as" and "Moscow"

Substitute: "which" for "what" in both underlinings

Substitute: "the descriptions" for "our knowledge"

Delete: "as"

Alternatively, we must look at representational functions in context. In context means analyzing what the individual components contribute to the containing representation in question. This interpretative setting serves as the basis for inferences about these elements (Danneberg, 2006). On this account, neither the Napoleon in Tolstoi's *War and Peace* nor the pig Napoleon in Orwell's *Animal Farm* denotes the real Napoleon. Instead, these two figures refer to our knowledge about the real Napoleon without being used to argue or certify any claim about Napoleon as a historical person. Therefore, these fictional Napoleons constitute an image of a denoting element, meaning they are not used to denote but refer to other items that are. The function of Napoleon as an element in *War and Peace* is determined by its role in the fictional world-story, not by its epistemic connection to historical sources (Barwich, 2013).

Substitute: "representations" for "they"

My second claim states that each form of representation is incomplete. Incomplete means they are limited and selective in their descriptions concerning real world properties. For example, a map only contains specific elements depending on its purpose (e.g., transportation maps give you information about the underground and bus system, but not about altitude). Notably, advocates of scientific pluralism present a similar argument when they concern scientific models. Models are necessarily limited in their content and scope. Their boundaries resonate with their epistemic objective: Are we aiming for realism, generality, or precision when using a model (Levins, 1966)? This methodological argument for pluralism derives from the ontological complexity of the physical world. To capture the multi-leveled, overlapping, and contingent features of nature, especially of biological entities and processes, we need a mosaic of models (Mitchell, 2003; Wimsatt, 2007).

Delete: comma after "argument"

Now, this has further consequences for our interpretation of such models. We sometimes look for information that is not explicit in a representation (for a more detailed argument, see Barwich, 2014). Sometimes this information is implicit. For example, reading a novel about two people falling in love, you possibly assume that these two individuals have a heart, a liver, and at least one functioning kidney, even if the author fails to mention this. It is because we know that people usually do not survive without these organs. We can make such explicit inferences even in fiction because any representation is somewhat "parasitic" on our common knowledge about the world and the use of language concepts (Searle, 1975; Eco, 1994). Therefore, unless stated explicitly, any word or concept in fictional discourse has the same meaning and implications as it has in non-fictional discourse.

Insert: "otherwise" after "explicitly"

What if we want to know about something that is not inferable in such an implicit sense? Some questions lead us to the limits of interpretation if they are not answerable by inferences based on the representation in question. For example, ask yourself, *how many children had Lady Macbeth?* There is no plain answer to this issue. Shakespeare's Lady Macbeth may not be of the nurturing kind, but she could have had children, even though none are mentioned explicitly in the play (Knights, 1933).[3] Likewise, we know that Captain James T. Kirk from the starship *Enterprise* is blond. However, is he a natural blond?

Correction: "blonde" instead of "blond"

The vanity of his character does not exclude the possibility that he colors his hair regularly. We will never know for sure as long as no conclusive information is provided. The same issue goes for the interpretation of scientific models. The chemical formulas of Berzelius, for example, give information about proportions but not about mechanical features of atoms, such as their size or shape (Klein, 2001). Nonetheless, in an ongoing experimental pursuit, such missing information is not left aside. It is addressed by amending the model or, alternatively, by looking for an answer through alternative models in the same research context. Identifying and clarifying such limits of models spurs further scientific inquiry.

Persistent failure to provide a better account of such missing information in a model can also lead to serious doubt about the model's referential grounds. Consider the case of phlogiston. Central to its ontology was the question of whether phlogiston is weightless, and the continuing inconsistencies in answering this issue were a reason for abandoning the substance (Kim, 2008; Barwich, 2013).

Substitute: "concept" instead of "substance"

The upshot is that identifying and testing the limits of scientific representations is a useful way to determine whether a particular inference is only a model-dependent, fictional truth. Overall, this section outlined two central characteristics of representations, namely hybridity and incompleteness, which further allow us to compare and distinguish fictional and non-fictional representations. The remainder of the chapter now elaborates on how we can use these two characteristics to make decisions about ambiguous elements in scientific modeling contexts. For this, I now look at how scientists make sense of indeterminate inferences when using X-ray crystallography for protein models. Indeed, X-ray crystallography is an excellent example to analyze the problem of representation, as we will see that crystallographers face several of the representational issues mentioned above.

Delete: "contexts"

3 The case of X-ray crystallography

Correction: "X-ray" instead of "X-Ray"

X-Ray crystallography is a principal method in biochemistry. It serves to determine the molecular structure of macromolecules such as proteins, and it works like this (Figure 4.1): Protein materials are prepared in specific detergents so that they form neat crystalline structures. These crystals are mounted on a goniometer (an instrument that allows the rotation of the inserted object). When placed in the goniometer, the crystals are shot with beams of X-rays. These X-rays scatter on the crystal surface and form a diffraction pattern, which is collected on an image plane or X-ray film. This diffraction data accounts for the electron density in the crystal structure. It serves as the basis to infer atom positions and, consequently, the molecular structure of the crystal (Drenth, 2007; Serdyuk, Zaccai, and Zaccai, 2007; Smyth and Martin, 2000).

Successful applications of this procedure are not without trickery and difficulties, of course. There is no simple mapping of the models onto the raw materials. Rather, instrumental access to macromolecules is facilitated through

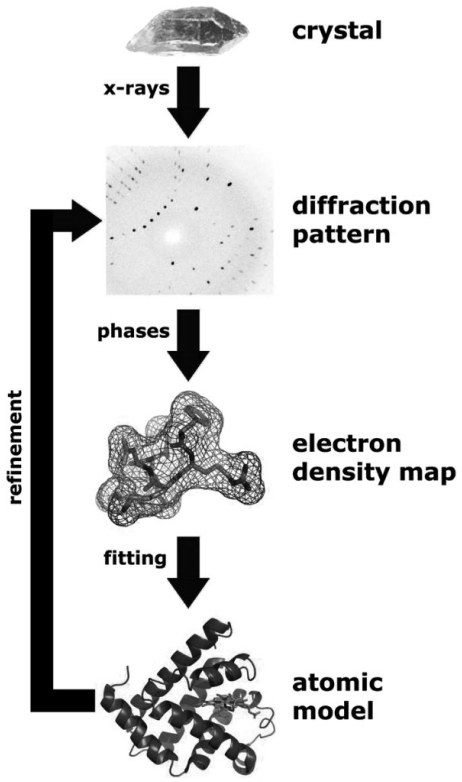

Figure 4.1 Basic steps involved in X-ray crystallography (image from Splettstoesser 2006). First, proteins are crystallized to form symmetric structures. These crystals are treated with X-ray beams to obtain diffraction data based on the electron density of the crystal structure. Diffraction patterns are further translated into electron-density maps from which to infer the position of atoms and the basic molecular structure of the protein. The resulting atomic model of the protein is refined by successive and iterative cycles of this procedure.

several steps of manipulating the materials to fit the model requirements. The foundation of protein models in X-ray crystallography is not so much a reconstruction but the very production of certain structures. The success of inferences based on this method is contingent on the multiple steps for bringing the materials in correspondence with the requirements of the model procedure. Each of these steps carries its ambiguities and modeling problems.

3.1 Distortions, or: making the materials fit the method

The essential material precondition for X-ray crystallography is symmetry. So, why do we need symmetric crystals in the first place? Creating symmetric structures from organic matter is a difficult issue, but essential for successful data collection.

To make crystals, you start out with proteins in a solution of high concentration. You purify them and slowly remove the water by putting the materials in a specific detergent. However, the crystallization of proteins can be a

daunting task, and it often involves a laborious trial-and-error procedure. "The magnitude of the problem can be understood when one considers the variables: the choice of precipitant, its concentration, the buffer, its pH, the protein concentration, the temperature, the crystallization technique, and the possible inclusion of additives (Smyth and Martin, 2000: 8)."

Some macromolecules form nice, symmetric crystalline structures. Others remain stuck in an unsymmetrical and flat formation. A particularly salient example for this are transmembrane proteins, such as the superfamily of G-protein-coupled receptors. G-protein-coupled receptors, or GPCRs, constitute the largest protein gene family in the mammalian genome, and are involved in several fundamental biological processes such as vision, olfaction, the regulation of immune responses, and the detection of neurotransmitters. However, their instability prevents these transmembrane proteins from building regular crystalline structures easily. The first success in getting the structure of an active ternary complex of a GPCR through X-ray crystallography happened only a few years ago (Snogerup-Linse, 2012; Kobilka, 2013).[4] Any attempt to crystallize proteins from the largest member of this protein superfamily, the olfactory receptors, has been unsuccessful to date (Crasto, 2009; Barwich, 2016).

These difficulties aside, regular applications of X-ray crystallography require three-dimensional symmetric crystals. Symmetry is indispensable for combining the series of diffraction images that are collected while the crystal is rotated. Symmetry allows applying mathematical interpretations to these diffraction patterns and to turn them into electron-density maps. The reason for this is because crystals facilitate the determination of "unit cells," those parts of the crystal that form its smallest repeating units (Figure 4.2). These unit cells act as a metric index to determine the dimensions of the overall crystal structure. They further relate the series of individual diffraction images to each other: "Computer programs for autoindexing do this by calculating a prediction of

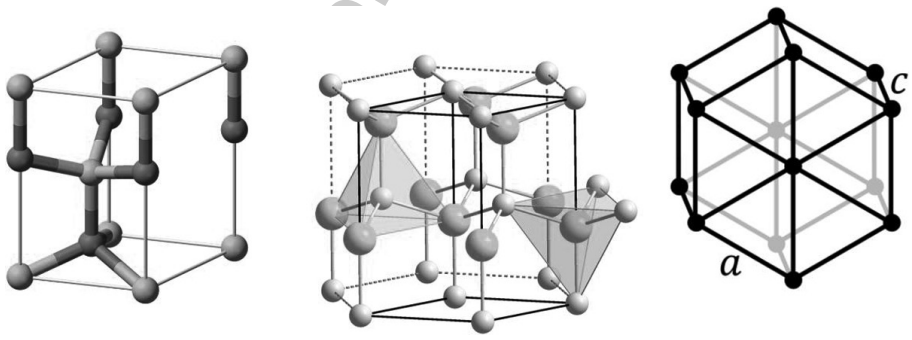

Figure 4.2 Schematic representation of the principle of unit cells in X-ray crystallography with the example of the Wurtzite crystal structure (a mineral structure). Unit cells are the smallest repeating unit (left) from which to infer the general structural dimensions of the crystal (middle) and its geometric arrangement of the crystal (right). Unit cells act as an index. Images taken from Benjahbmm27, 2007 (left); Solid State, 2008 (middle); Mayer, 2007 (right).

what the diffraction image will look like from the cell dimensions and orientation, then attempting to fit the real image with the predicted one (Smyth and Martin, 2000: 12)."

This is why you need symmetric crystals for the method to work. Data from flat or unsymmetrical crystals are incomplete, distorted, and impossible to accommodate with the mathematical tools (Smyth and Martin, 2000). Distorted and indeterminate data pose significant limits for legitimate inferences to the molecular structure and constitution of macromolecules, such as proteins. However, even obtaining "good enough" diffraction data has its difficulties and limits. As mentioned earlier, some proteins are harder to crystallize than others, such as transmembrane receptors. Furthermore, proteins, when treated with X-rays, disintegrate quickly and, as a result, the collected diffraction data can be incomplete or insufficient. Concerning the gradual disintegration of proteins during the procedure, the first diffraction image is usually of the best quality.

This issue of indeterminate data is more than a matter of temporary technological concern. It marks the methodological dependence of experimental research on the available instrumental tools and the structural aspects of the material to which they are responsive. Successful inferences to the molecular dimension of proteins do not ground in an essential trait or structure of the raw materials per se. On the contrary, proteins are irregular and dynamic when untreated. Instead, methods such as X-ray crystallography rely on artificially-produced features, such as symmetry. That said, the production of symmetric crystals is not the only methodological requirement for successful model-building through X-ray crystallography.

3.2 Incomplete information in model-based inferences

The next basic step is the analysis of diffraction data and their transformation into electron-density maps. What we see in recordings of diffraction patterns is a distribution of electron density (Figure 4.3). The concentrated rings of spots are the result of diffracted X-rays that are emitted from the crystal and collected on an imaging plane. These spots indicate electron clouds. Now, how do we get to the molecular structure of proteins from a measure of electron density? Meaning, how do we infer atom positions from these electron spots?

This is a matter of data-processing. The analysis of diffraction patterns rests on the distinction between meaningful diffraction data and mere background noise. This distinction is crucial for further calculations and to transform the diffraction data into readable electron density maps (Figure 4.4). These density maps then allow us to infer atom positions and the molecular organization of the protein. Notably, in the early applications of X-ray crystallography, this was the very problem. What is the noise to data ratio? Moreover, what do we see when we look at the diffraction data?

X-ray beams come in waves, and waves have different phases. However, we can only record the overall intensity of beams on the image planes. Thus, the

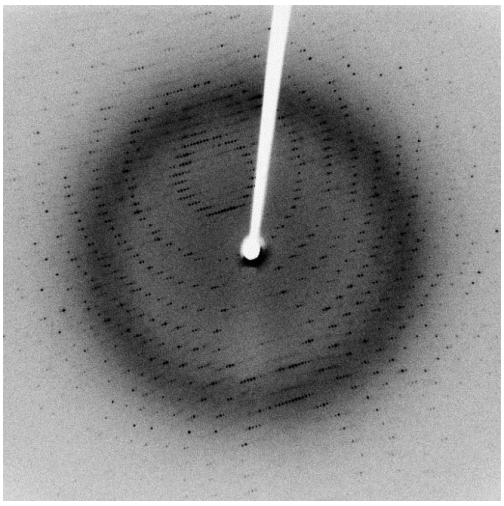

Figure 4.3 X-ray diffraction pattern (image from Dahl, 2006). Concentrated rings of spots are the result of the diffracted X-rays, which are emitted from the crystal and collected on an imaging plate. Spots indicate electron clouds.

Correction: "plane" instead of "plate"

Fig. 4.4 Representation of a protein crystal structure where atoms and molecular units are built into an electron density map. (Image from Bassophile 2007.)

recordings do not tell us anything about the particular phase the waves are in when hitting the plate. Are these waves in sync (in phase) or not (out of phase)? To be sure, different phases result in different intensities of the recorded spots. However, to understand what these diffraction spots represent, and to infer atom positions, we need to know in which phase the X-ray beams are when they hit the plate (Figure 4.5). How does one get access to this information?

Correction: "plane" instead of "plate" in both underlinings

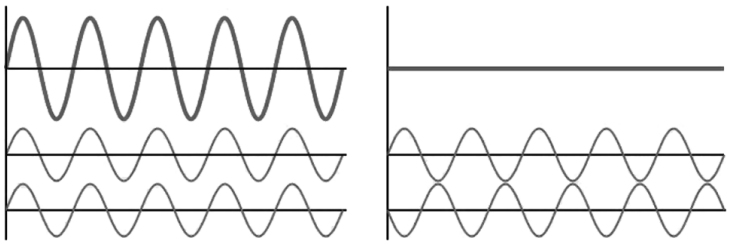

Figure 4.5 Representation of different wave phases: in sync (left; from Haade, Wjh31) and out of sync (right; Quibik, 2010); the diffraction spots from in sync phases are stronger than those from out of sync waves.

The mere diffraction patterns do not allow for inferences to the wave phases, and the issue is known as the "phase problem."

Solving the phase problem was an ingenious piece of work by the molecular biologist Max Perutz at Cambridge, 1962 Nobel Laureate in Chemistry (Perutz, 1942; 1962). Perutz wanted to solve the molecular structure of hemoglobin, a complex protein. He saw himself confronted with the missing phase information, resulting in a lack of representational stability and a strong ambiguity in data recordings. The solution to this query he obtained by adopting a method previously used by of one of his colleagues, J. M. Robertson at Glasgow University: isomorphous replacement (Pietzsch, 2016).

Isomorphous replacement works by soaking the protein in a heavy atom solution, such as mercury or platinum. The protein incorporates one or more of these heavy atoms, but this does not change its spatial configuration. Nonetheless, the incorporation of heavy atoms modifies the diffraction patterns, since the "heavier" protein now contains more electrons. Therefore, we end up with two sets of data: the diffraction patterns from the original protein and the patterns of the protein containing heavy atoms. Comparing these data sets, any differences in recorded intensities are due in large part to the presence of heavier atoms. From this we can derive the missing phase information. The subtle differences in intensities between the data sets allow us to find the position of the heavier atoms. These atoms then act as reference points to infer the phases (Smyth and Martin, 2000; Cowtan, 2003; Pietzsch, 2016).

The point that the resolution of the phase problem illustrates is that the interpretation of structural correspondence is a product of material manipulations. Therefore, inferences derived from these structures must be judged carefully against the model-building requirements. What we see through such a detailed perspective on the process of model-building is again that the structural correspondence between a model and the materials reflects the methodological conditions of the procedure, and not necessarily essential, raw features of the materials. To be sure, presumptions about structural features are an integral part of modeling. Nonetheless, they cannot provide independent criteria for the evaluation of model-based inferences and the referential capacity of a scientific presentation.

Correction: "representation" instead of "presentation"

The third step is building the final protein model. For this, the quality of the electron-density map is crucial to determine the arrangement of molecular units for a three-dimensional model. Electron maps provide the basic outline wherein the protein structures are built (Figure 4.4). The higher the resolution of these maps, the less ambiguous is the identification of the relevant molecular units and subunits. The derived image presents us with a three-dimensional model of the protein structure. Lastly, the outputs are formatted and placed in the protein data bank (PDB).[5] An integral part of this concluding step is the continuous refinement of the data analysis. Specifically designed molecular visualization programs facilitate the interpretation of structural features of the molecule, such as bond lengths and angles (Smyth and Martin, 2000).

4 Integrative pluralism as a representational requirement in scientific practice

At this point, let us come back to the general issue of representation and the epistemic role we assign to scientific representations. In examining the details of X-ray crystallography, I deliberately avoided discussing how the problems in this procedure resonate with the debate about fictionalization strategies in scientific representation. My reasons for this was because I wanted to focus in more detail on how this method works from the practitioners' perspective before connecting its modeling steps to the philosophical discussion at hand.

To bring the different modeling steps together for further analysis, so far we have seen that multiple factors play a major role when building a protein model through X-ray crystallography. These factors concern the experimental manageability of the research materials, the generation of a sufficient range of data, the availability of appropriate methods to translate the diffraction patterns into electron-density maps, and also the introduction of data-processing techniques, such as molecular visualization programs. All these factors are involved in a series of material and conceptual operations that shape and result in the final receptor model. In brief, these sequential steps comprise of:

Delete: "of"

- first, the material transformation of flexible proteins into stable and rigid crystal structures (fulfilling the requirement of symmetry);
- second, the material inscription from crystallized protein structure to diffraction data (relying on Bragg's model, or Bragg's law);[6]
- third, the translation of diffraction data into electron density maps through the Fourier transform);[7]
- fourth, the subsequent inferences from the electron density map to a three-dimensional model of the protein (using computational programs to calculate and visualize the positions and relations of atoms as inferred from the electron clouds).[8]

The detailed dissection of modeling steps in X-ray crystallography revealed a chain of mediation between the raw materials and the final model. This chain

made explicit the several requirements by which we generate "structures" as a meaningful model ingredient. Although the successive modeling steps do not follow logically from each other, the manifestation of these steps is informed by, and grounded in, the inferences and structural correspondences established through their preceding ones. On this account, an analysis of scientific representations based on dyadic and abstracted notions such as "structural similarity" or "resemblance" between a model and its target system must look too simplistic to the practitioner to be of practical utility. To her, it matters more to suggest a heuristic perspective on how to deal with the ambiguous components in a modeling procedure, such as the phase problem. For this, a more detailed understanding of the method is necessary.

Every model has a history that determines its construction and use and, in turn, its potential and limits for inference to the properties of the investigated materials. The models we derive from scientific methods are, therefore, not freely floating objects that correspond to the world through some depiction of abstracted features. Instead, these abstracted features are a representation of the steps that link specific structural assumptions to the materials in a series of manipulations. From this view, any evaluation of a model and the inferences drawn from it must be judged against the background knowledge on which the modeling procedure is based. In the case of crystallography, to assess how the derived protein structures do (or do not) account for protein properties and behavior, a mere look at the finalized end product is insufficient.

The fictionalization strategies we encountered here included: (1) distortions (forcing proteins into symmetric crystals); (2) idealizations (predictions of crystal structure through autoindexing of unit cells to align diffraction planes); (3) indeterminate data (disintegrating crystals); and (4) the underdetermination of models by data (recordings of electron spots as part of the phase problem). Indeed, it appears miraculous that this procedure could become one of the most reliable and standard methods in structural biology and biochemistry at all!

That said, these issues demonstrate the relevance of my two earlier hypotheses, namely the hybridity and the incompleteness of representations. Regarding the first hypothesis, these fictionalization strategies embody the hybridity of crystallography models where certain elements do not strictly denote features of proteins but reflect the requirements of the modeling method. Concerning the second hypothesis, the incompleteness of representations, we may ask: How are the ambiguous model elements dealt with in X-ray crystallography? Accordingly, how do the practitioners answer those questions about the materials that exceeded the available model structures? The kinds of issues that exceed a method usually arise from the particular purpose a method is designed to serve.

Correction: "exceed" instead of "exceeded"

The target of protein models, constituting the Holy Grail of modern biochemistry, is to solve the relation between a protein's structure and its function. Thus far, however, drug target studies and approaches to rational drug design operate on a labor-intensive trial and error basis (Drews, 2000). A pivotal reason for the lack of principles after which to model structure-function relations

Substitution: "motifs" instead of "sequences"

in proteins is the problem of protein-folding. After decades of elaborating on protein structures, it became apparent that there is no straightforward way to make predictions from the amino acid sequences to the three-dimensional configuration of proteins.[9] Proteins often fold irregularly. However, the folding process fundamentally determines protein function. As a result, model-based inferences to ligand-binding in proteins require three-dimensional snapshots of proteins. This issue also explains why X-ray crystallographers are still in business (Mitchell and Gronenborn, 2015).

Insert: "the" between "of" and "proteins"

Nevertheless, the irregularity between protein structure and function not only explains the continuing demand for crystallography. It also presents the limits of this procedure. A major concern for biologists is precisely what the protein models from X-ray crystallography do not show and the questions they cannot answer.

Insert: "based on" between "is" and "a highly"

Proteins are extremely specific in behavior and function. Their functioning, such as in molecular recognition, is a highly dynamic mechanism, involving several steps of conformational changes of proteins. Modern biology has therefore abandoned the lock-and-key metaphor. Instead, biologists refer to models of "induced" (Koshland, 1995) and "selected fit" (Monod, Wyman, and Changeux, 1965; Changeux, 2013). Here, proteins are not rigid but dynamic entities. They undergo multiple changes in their conformation. Several questions arise from such models of conformational changes, such as: Are these changes induced by the ligand, or do these changes take place in the absence of a ligand? (For more details on these mechanisms, see Barwich, 2016.)

Delete: comma after "meachanisms"

Substituion: "the" instead of "this"

X-ray crystallography is unsuitable for determining the sequences of flexible conformational changes in protein-binding. It is also unfit to make inferences about the specific causal role of the ligand in this recognition mechanism. Yet, these are two essential issues for understanding protein behavior. From this view, the strength of this method is also the source of its ambiguity: How can we make inferences about protein behavior from the rigid representations of protein structure?

Insert: "of proteins" after "binding behavior"

Substitution: "the" instead of "such"

Despite its overall success, concern about the limits of X-ray crystallography has always been vocal. A source of such concern often goes back to the inevitable incompleteness of representations. We have seen this to be the case for inferences about the binding behavior. Additionally, such characteristic of incompleteness is inevitably part of the model-building procedure, not only of the final product. For example, the big issue in the earlier applications of this method was the ambiguity of what the image planes show. When Perutz presented his diffraction patterns, it was Francis Crick, his student, who ambushed him severely in a public talk regarding the ambiguity of this data in light of the unresolved phase problem (Pietzsch, 2016). The issue was resolved after an additional procedure was implemented into the model. The data from isomorphic replacement resonated precisely with the structural requirements of the method, and this correspondence then allowed determining of the unknown phases. To be sure, once the phase problem was solved, other limits of the procedure were rendered visible.

Correction: "a determination of" instead of "determining of"

Delete: "Nonetheless," And start sentence with "Scientists"

Nonetheless, scientists rarely judge the interpretation of a model in isolation. Additional procedures are used in parallel with a model to determine which parts of its inferences can be corroborated and complemented by other means or may constitute a potential artifact. Of particular importance is a satisfactory independence of some modeling parameters in the corroborating procedure. Of course, the issue then is to ascertain that the different procedures account for the same phenomenon and are coordinated with each other (Barwich and Chang, 2015).

For example, complementary to X-ray crystallography, another way to build protein models is nuclear magnetic resonance (NMR) spectroscopy. One of the limits of crystallography is the requirement of isolating the protein from its cellular environment and studying it through a crystalline state. In NMR, proteins are used in a soluble state. "Furthermore, X-ray crystallography and NMR target different atomic features of a protein: X-ray crystallography relies on the scattering of X-rays by the electrons, while in NMR, interactions of nuclear spins with a magnetic field are explored (Mitchell and Gronenborn, 2015; 15)." In comparison, each method results in mediated representations of proteins outside their cellular environment. Still, both approaches count as reliable representational sources for inferences about protein behavior in vitro.

Overall, the difficulty of indeterminate inferences – meaning inferences that may not represent a physical entity but constitute a model artifact – are addressed through a comparison of data through different procedures. Such comparisons can be an indicator of inaccuracies and a reflection of a method's biases. To correct these biases requires a deeper understanding of the structural requirements that fit the materials to the method in the model-building procedure.

Correction: "representations" instead of "representation"

In closing, we must understand inferences to information that is not explicitly given in the model as means to determine how this model is linked to other methods in an experimental context. In contrast to fiction, interpretations of ambiguous elements and incomplete representation are how science proceeds. From this perspective, the epistemic distinction between science and fiction is not one inherent in a representational structure but in our way of dealing with such limits. Therefore, the integration of a scientific model with other methods ensures that its inferences are testable as being non-fictional (i.e., representative of a phenomenon) or as fictional (i.e., model-dependent and an artifact of the procedure). As science is ongoing, a scientific representation cannot be final. Therefore, its representational function and limits must be probed through its integration in an operating model context.

Insert: "and ongoing" between "operating" and "model"

Notes

1 By representations, I mean all forms of public description and depiction, whether these are linguistic, algebraic, symbolic, pictorial, and so on.

2 Silogens are hypothetical entities in nanomechanical models. They are used to calculate silicon fractures and to combine theoretical assumptions from quantum mechanics and classical molecular dynamics. Silogens do not refer to real atoms. Instead, they are algo-

Delete: "to'

rithmic combinations of the properties of two different entities in the modeling procedure, namely silicon and hydrogen (Winsberg, 2009). Notably, my interpretation of the epistemic status of silogens differs from Winsberg's (Barwich, 2014).

Substitute: "the" instead of "this"

3 In a famous essay, entitled 'How Many Children Had Lady Macbeth?', the Shakespeare scholar L. C. Knights (1933) asked about the inferential limits in fictional works. This essay argued against the tendency of excessive over-interpretations in literature studies at this time.

4 Brian Kobilka received the 2012 Nobel Prize in Chemistry for this achievement. He shared this nomination with his former mentor Robert Lefkowitz, who received the award for his crucial work on the ß-adrenergic receptors and the general workings of GPCRs (Snogerup-Linse, 2012; Clark, 2013).

5 Website of the PDB: http://www.wwpdb.org

Correction: "Lawrence" instead of "Laurence"

6 Cambridge physicist William Laurence Bragg and his father William Henry Bragg, professor of mathematics at the University of Adelaide, determined the angles in which X-rays scattered from a crystal lattice. Together they were awarded the 1915 Nobel Prize in Physics. Lawrence Bragg was only 25 at the time and has been the youngest Nobel Prize winner in Physics to date (Nobel Media, 2016).

Delete: comma after "physicist"

7 The Fourier transform allows calculating the frequencies that make up a signal (here, the X-ray beam). The Fourier transform is named after the nineteenth-century French mathematician and physicist, Joseph Fourier.

8 Examples for such molecular visualization programs are RasMOL and MOLMOL.

9 When we speak of protein structure, we must differentiate between four levels: primary structure (amino acid sequences), secondary structure (regular subunits of a protein such as helix domains), and tertiary structure (three-dimensional protein folding), and the quaternary structure (three-dimensional structure of multi-subunit proteins).

Bibliography

Bib contains minor corrections (thanks to an Endnote mess up)

Ankeny, R. A. and S. Leonelli. (2011). 'What's So Special about Model Organisms?'. *Studies in History and Philosophy of Science Part A*, 42 (2): 313–23.

Barnes, B., D. Bloor and J. Henry. (1996). *Scientific Knowledge: A Sociological Analysis*. Chicago: University of Chicago Press.

Barwich, A.-S. (2013). 'Science and Fiction: Analysing the Concept of Fiction in Science and Its Limits'. *Journal for General Philosophy of Science*, 44 (2):357–73.

Insert pages: 291-309

Barwich, A.-S. (2014). 'Fiction in Science? Exploring the Reality of Theoretical Entities', in G. Bonino, G. Jesson and J. Cumpa (eds.) *Defending Realism: Ontological and Epistemological Investigations*. Boston, Berlin and Munich: de Gruyter.

Barwich, A.-S. (2016). 'What Is So Special about Smell? Olfaction as a Model System in Neurobiology'. *Postgraduate Medical Journal*, 92: 27–33.

Barwich, A.-S. and H. Chang. (2015). 'Sensory Measurements: Coordination and Standardization'. *Biological Theory*, 10 (3):200–211.

Bassophile. 2007. 'Eden.png.' *Wikipedia*. Accessed 26 June. https://en.wikipedia.org/wiki/File:Eden.png.

Benjah-bmm27. 2007. 'Wurtzite-unit-cell-3D-balls.png.' *Wikipedia*. Accessed 26 June. https://en.wikipedia.org/wiki/File:Wurtzite-unit-cell-3D-balls.png.

Boumans, M. (1999). 'Built-in Justification', in M. Morgan and M. Morrison (eds.) *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Chang, H. (2012). *Is Water H2O?: Evidence, Realism and Pluralism*. Vol. 293, New York: Springer.

Changeux, J-P. (2013). '50 years of Allosteric Interactions: The Twists and Turns of the Models'. *Nature Reviews Molecular Cell Biology*, 14: 819–29.

Clark, R. B. (2013). 'Profile of Brian K. Kobilka and Robert J. Lefkowitz, 2012 Nobel Laureates in Chemistry'. *Proceedings of the National Academy of Sciences*, 110 (14): 5274–5.

Cowtan, K. (2003). 'Phase Problem in X-ray Crystallography, and Its Solution'. *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons Ltd.

Crasto, C. J. (2009). 'Computational Biology of Olfactory Receptors'. *Curr Bioinform,* 4 (1):8–15.

Dahl, J. (2006). 'X-ray Diffraction Pattern 3clpro.jpg'. *Wikipedia*. Accessed 23 June. https://en.wikipedia.org/wiki/File:X-ray_diffraction_pattern_3clpro.jpg.

Comment below

Danneberg, L. (2006). 'Weder Tränen noch Logik. Über die Zugänglichkeit fiktionaler Welten'. *Forschungsstelle Historische Epistemologie und Hermeneutik*. Accessed 31 February. http://fheh.org/images/fheh/material/danneberg-fiktion.pdf.

Deutsch, D. (2011). *The Beginning of Infinity: Explanations that Transform the World*. London: Penguin UK.

Drenth, J. (2007). *Principles of Protein X-Ray Crystallography*. New York: Springer Science & Business Media.

Drews, J. (2000). 'Drug Discovery: A Historical Perspective'. *Science,* 287 (5460): 1960–4.

Eco, U. (1994). *Six Walks in the Fictional Woods*. Cambridge, MA: Harvard University Press.

Frigg, R. (2002). 'Models and Representation. Why Structures Are Not Enough'. *Measurement in Physics and Economics Project Discussion Paper Series,* London School of Economics. Accessed 31 February. http://www.romanfrigg.org/writings/Models_and_Representation.pdf.

Frigg, R. (2010a). 'Fiction in Science', in J. Woods (ed.) *Fiction and Models: New Essays*. 247–87. Munich: Philosophia Verlag.

Frigg, R. (2010b). 'Models and Fiction'. *Synthese*, 172 (2):251–68.

Goodman, N. (1968). *Languages of Art: An Approach to a Theory of Symbols*, 2nd edition. Indianapolis: The Bobbs-Merrill Company.

Haade, Wjh31, and Quibik. (2010). 'Interference of two waves.svg'. *Wikipedia*. Accessed 26 June. https://en.wikipedia.org/wiki/File:Interference_of_two_waves.svg.

Hesse, M. B. (1966). *Models and Analogies in Science*, *Vol.* 36. Notre Dame: University of Notre Dame Press.

Holyoak, K. J. and P. Thagard. (1996). *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: MIT Press.

Kim, M.G. (2008). 'The "Instrumental" Reality of Phlogiston'. *Hyle*, 14 (1):27–51.

Klein, U. (2001). 'Berzelian Formulas as Paper Tools in Early Nineteenth-Century Chemistry'. *Foundations of Chemistry*, 3 (1):7–32.

Knights, L. C. (1933). *How Many Children Had Lady Macbeth?: An Essay in the Theory and Practice of Shakespeare Criticism*. Cambridge: The Minority Press.

Knuuttila, T. (2005). 'Models, Representation, and Mediation.' *Philosophy of Science*, 72 (5):1260–71.

Knuuttila, T., and A. Voutilainen. (2003). 'A Parser as an Epistemic Artifact: A Material View on Models'. *Philosophy of Science,* 70 (5):1484–95.

Kobilka, B. (2013). 'The Structural Basis of G-Protein-Coupled Receptor Signaling (Nobel Lecture)'. *Angewandte Chemie International Edition,* 52 (25):6380–8.

Koshland, D. E. (1995). 'The Key-Lock Theory and the Induced Fit Theory'. *Angewandte Chemie International Edition in English,* 33 (23–24):2375–8.

Correction:

Danneberg, L. (2006). 'Weder Tränen noch Logik. Über die Zugänglichkeit fiktionaler Welten', in: U. Klein, K. Mellmann and S. Metzger (eds.), Heuristiken der Literaturwissenschaft. Einladung zu disziplinexternen Perspektiven auf Literatur, Paderborn, 35–83.

Levins, R. (1966). 'The Strategy of Model Building in Population Biology'. *American Scientist,* 54 (4):421–31.

Longino, H. E. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Mayer, D. (2007). 'Hexagonal lattice.svg'. *Wikipedia*. Accessed 26 June. https://en.wikipedia.org/wiki/File:Hexagonal_lattice.svg.

Mitchell, S. D. (2003). *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.

Mitchell, S. D. and A. M. Gronenborn. (2015). 'After Fifty Years, Why Are Protein X-ray Crystallographers Still in Business?' Paper submitted to *The British Journal for the Philosophy of Science.*

Monod, J., W. Jeffries and J-P. Changeux. (1965). 'On the Nature of Allosteric Transitions: A Plausible Model'. *Journal of Molecular Biology,* 12 (1):88–118.

Morgan, M., and M. Morrison (eds.). (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Nersessian, N. (2010). *Creating Scientific Concepts*. Cambridge,MA: MIT Press.

Nobel Media. (2016). 'Facts on the Nobel Prizes in Physics'. Nobelprize.org. Accessed 26 June, 2016. http://www.nobelprize.org/nobel_prizes/facts/physics/.

Perutz, M. F. (1942). 'X-ray Analysis of Haemoglobin'. *Nature,* 149:491–4.

Perutz, M. F. (1962). 'Nobel Lecture: X-ray Analysis of Haemoglobin'. *Nobelprize.org*. Accessed June 26. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1962/perutz-lecture.html.

Pietzsch, J. (2016). 'The Nobel Prize in Chemistry 1962 – Perspectives'. *Nobelprize.org*. Accessed June 26. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1962/perspectives.html.

Rheinberger, H-J. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford: Stanford University Press.

Ryle, G. (1933). 'Imaginary Objects'. *Proceedings of the Aristotelian Society Suppl.,* 2:18–43.

Searle, J. R. (1975). 'The Logical Status of Fictional Discourse.' *New Literary History: On Narrative and Narratives,* 6 (2):319–32.

Searle, J. R. (2010). *Making the Social World. The Structure of Human Civilization*. Oxford: Oxford University Press.

Serdyuk, I. N., N. R. Zaccai, and J. Zaccai. (2007). *Methods in Molecular Biophysics: Structure, Dynamics, Function*. Cambridge, MA: Cambridge University Press.

Smyth, M. S. and J. H. J. Martin. (2000). 'X Ray Crystallography'. *Molecular Pathology,* 53 (1):8–14.

Snogerup-Linse, S. (2012). 'Studies of G-Protein Coupled Receptors. The Nobel Prize in Chemistry 2012. Award Ceremony Speech'. *The Royal Swedish Academy of Sciences*. Accessed 25 September 2014. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2012/advanced-chemistryprize2012.pdf.

Solid State. (2008). 'Wurtzite polyhedra.png'. *Wikipedia*. Accessed 26 June. https://en.wikipedia.org/wiki/File:Wurtzite_polyhedra.png.

Splettstoesser, T. (2006). 'X ray diffraction.png'. *Wikipedia*. Accessed 23 June. https://commons.wikimedia.org/wiki/File:X_ray_diffraction.png.

Suárez, M. (2004). 'An Inferential Conception of Scientific Representation'. *Philosophy of Science,* 71: 767–79.

Suárez, M. (ed.) (2009). *Fiction in Science: Philosophical Essays on Modelling and Idealization*. New York: Routledge.

Comment below

Delete: "Paper submitted"

Correct: The British Journal for the Philosophy of Science, http://doi.org/10.1093/bjps/axv051

Toon, A. (2012). *Models as Make-Believe: Imagination, Fiction and Scientific Representation*. London: Palgrave Macmillan.

van Fraassen, B. C. (2010). 'Scientific Representation: Paradoxes of Perspective'. *Analysis*, 70 (3):511–14.

Walton, K. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard: Harvard University Press.

Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard: Harvard University Press.

Winsberg, E. (2009). 'A Function for Fictions: Expanding the Scope of Science', in M. Suárez (ed.) *Fiction in Science. Philosophical Essays on Modelling and Idealisation*. London: Routledge.

Comment below

Correction: "Cambridge, MA" instead of "Harvard" in both underlinings

THANK YOU, STEVEN!!