Explaining with Simulations: Why Visual Representations Matter

Julie Jebeile
Institut supérieur de philosophie, Université catholique de Louvain

Computer simulations are often expected to provide explanations about target phenomena. However there is a gap between the simulation outputs and the underlying model, which prevents users finding the relevant explanatory components within the model. I contend that visual representations which adequately display the simulation outputs can nevertheless be used to get explanations. In order to do so, I elaborate on the way graphs and pictures can help one to explain the behavior of a flow past a cylinder. I then specify the reasons that make more generally visual representations particularly suitable for explanatory tasks in a computer-assisted context.

1. Introduction

Mathematical models are often expected to provide not only predictions about the phenomenon that they represent, but also explanations (see, e.g., Bokulich 2009, 2011; Heidelberger 2006; Morrison 2009). These explanations are answers to why-questions and particularly answers to why the predicted phenomenon should occur. For instance, models can be used to calculate when the next total solar eclipse will happen, and then to explain why it will take place on July 2, 2019. In this regard we can obtain explanations from a model if we can solve the model equations which govern the phenomenon under study. But some equations have no explicit solution or are too complicated to solve. In these cases it is difficult for a human mind to derive the solutions from the model. This difficulty reveals a "complexity barrier" beyond which the explanation of the phenomenon is not accessible without computer assistance (Lenhard

I am very grateful to Anouk Barberousse for her helpful comments on a previous version of this paper.

Perspectives on Science 2018, vol. 26, no. 2
 doi:10.1162/POSC_a_00273

2006). However, computer simulations are now used by scientists to overcome this barrier. They notably enable one to solve non-analytical and complex equations. Besides, the speed of the computational devices allows one to perform calculations faster than human beings. We might therefore expect that computer simulations increase our possibilities to get scientific explanations.

However, answering why-questions with models involves searching for the relevant explanatory components within the model. And yet, in order to find these components, it is not enough for the scientist to know the simulation model, i.e., the model equations, the initial conditions, and the boundary conditions. The search for relevance is made difficult because there is a gap between the simulation outputs and the model. This gap is largely due to the fact that computer simulations are epistemically opaque (Humphreys 2004); mainly they run too fast for one to follow the computational processes in detail and, even if it was possible to slow down the simulation, the simulation would still be too long to be cognitively grasped by a human mind.

Given the gap between the simulation outputs and the underlying model, shall we conclude that simulations cannot provide any explanation at all? This conclusion would be obviously wrong since scientists often use simulations for a very explanatory aim. A recent response claims that a possibility for overcoming epistemic opacity can be found in developing meta-models like those designed by economists (Lehtinen and Kuorikoski 2007; Kuorikoski 2011). But usually, there is no such possibility when we want to investigate complex systems, e.g., turbulent flows, spin glasses, population genetic systems, or stock markets. What does remain when all that we have for describing target phenomena are simulation models?

In this paper, I answer that, even though there is a gap between the simulation outputs and the underlying model, appropriate visual representations—e.g., graphs, diagrams, pictures, maps, and films—can be used to obtain explanations; the underlying assumption, though, is that the simulation model must at least correctly represent the phenomenon in order to explain. In arguing for that, I first develop on why there is a gap between the simulation outputs and the underlying model. I then elaborate on the way graphs and pictures can help one to explain the behavior of a flow past a cylinder. Lastly, I specify the reasons that make more generally visual representations particularly suitable for explanatory tasks in a computer-assisted context.

2. Gap between Simulation Outputs and Model

In this section, I will develop on why there is a gap between the simulation outputs and the underlying model. First let me put forward a few

terminological proposals. What I call a model contains theoretical principles as well as simplifying assumptions. The simulation model—that I also call the underlying model—contains theoretical principles, simplifying assumptions, and mathematical approximations due to the numerical scheme required for calculations on computers. The program, written in a computer language, contains the algorithm which describes how to process calculations from the simulation model. "Computer simulations" designate these calculations. In this paper, this expression will not be used to refer to the simulated phenomenon on the computer screen as it is sometimes done in the literature.

The gap between the simulation outputs and the underlying model is a philosophical concept that can be only considered as valuable if one takes the perspective of a user who wants to explain a phenomenon with a simulation. (Otherwise, of course, there is *stricto sensu* no gap in that the simulation does connect the simulation outputs with the underlying model). Thus let me characterize the explanatory task of a user before showing why there is a gap.

Explaining with a mathematical model is about answering why-questions about the target system that the model represents. It requires searching for relevant explanatory components within the model. The search for relevance in explanations crucially matters and this has been well recognized in several accounts of explanation (e.g., in Woodward 2003; Salmon 1998; Batterman 2002—even if they have their own definitions of the term).

For the purpose of searching relevance, the user needs at least to know the content of the model. In the case of simulation models, the user may gain this knowledge because she conceived and/or wrote and/or carefully read the program.[1] That said, knowing the model is a necessary but not a sufficient condition for gaining explanations in the case of simulation models. The reason is that there is a gap between the simulation outputs and the underlying model. Let me develop on this point.

The gap is a relative concept whose assessment depends on the adopted account of how to find relevant explanatory components within the underlying model. Thus, depending on the adopted account, there are potentially several kinds of gap that hinder the search for relevant explanatory components within the model. I will present three kinds of gap here.

1. It might seem pointless to emphasize this but actually users sometimes do not have the occasion to carefully read the computer program because of labor division. In a research team, people are typically divided into a group of developers and a group of users. Programmers mainly focus on the implementation (i.e., the writing of the algorithms) and on the verification of simulations. While users apply simulations on concrete physical cases: they prepare the input files that contain the initial conditions and the boundary conditions, execute the computer program and post-process the simulation outputs. Thus users might not have access to all the details in the computer program.

First one may expect to reach relevance as soon as one has an analytic understanding of the underlying model. One has an analytic understanding of a model if one is able to tell how the simulation outputs result from the interaction of the different model components (see Frisch 2015). The more simplified and idealized a model is, the easier it is to have this analytic understanding and thus to identify the relevant explanatory components among physical principles, initial and boundary conditions. Because, in simple models, equations simply and explicitly express the relations of dependence between variables of interest, the user may analytically penetrate these relations and in this way, may mentally make the relation between the empirical consequences from the model and the model itself. However, in simulation models, there are many variables and the relations of dependence are often non-linear and complicated. It therefore seems unclear how the different model components interact with each other and thus how simulation outputs are obtained. In this sense, there is a gap that one cannot fill based on the sole knowledge of the underlying model. In the case of climate models, for example, the lack of analytic understanding has at least three sources. A first source is that the climate system is a complex system composed of various heterogeneous components (e.g., general circulation of the atmosphere, cloud formation, sea and iceberg dynamics, vegetation effect) which interact with each other in a complicated and sometimes non-linear way. This is also due to the fact that climate models are characterized by their entrenchment (Lenhard and Winsberg 2010; Winsberg 2012): the track of the design choices made during their creation is not available in that these choices have been made by different individuals at different times. Thus scientists are not always able to justify these choices. Lastly, the components interact in the simulations in a way described in a "kludge," which is "an inelegant, 'botched together' piece of program; something functional but somehow messy and unsatisfying" (Clark 1987, p. 278; quoted in Lenhard and Winsberg 2010, p. 257).

Second, despite lacking analytic understanding of the underlying model, one may expect to find relevance by reconstructing the relation between the simulation outputs and the underlying model. Here the simulation is thought to be the only thing that stands for the demonstration of the simulation outputs from the model components. Finding relevance requires at least following the series of logical and mathematical operations that simulations are, before even trying to provide a short demonstration of how the simulation outputs were obtained. This series is generally not edited in practice, but let us consider that it can be made available for the sake of the argument. On this account, the user should go all over the simulation in *extenso* and survey every step of the computation, which may help her to track how the simulation outputs were obtained from the model components.

However, there is a gap here—in another sense than the previous one—which is due to the fact that computer simulations are epistemically opaque. On opacity, Humphreys writes:

> In many computer simulations, the dynamic relationship between the initial and final states of the core simulation is epistemically opaque because most steps in the process are not open to direct inspection and verification. This opacity can result in a loss of understanding because in most traditional static models our understanding is based upon the ability to decompose the process between model inputs and outputs into modular steps, each of which is methodologically acceptable both individually and in combination with the others. (Humphreys 2004, p. 148)

Opacity thus prevents the user comprehending the whole detailed process by which the simulation produces the connection between the model assumptions and the simulation outputs; it has at least two sources (see Humphreys 2004).

The first source is that simulations run so fast that no human brain could follow or survey the computational processes in detail. For this same reason, computationally assisted proofs of mathematical theorems, such as the four-color theorem, are often controversial (McEvoy 2008). Even in the case where their speed was reduced and adapted to the cognitive skills of the user so that she could follow the simulation unfolding,[2] she would need a lot of time, due to the large number of calculation steps, to follow the simulation entirely and she would not be able to cognitively grasp it anyway. Each computational step is understandable, but it is not possible to master the simulation in *extenso*.

The second source of opacity concerns what Stephen Wolfram has called "computationally irreducible processes" (Wolfram 1985, pp. 737–50). For Wolfram (1985), the evolution of a physical system may be calculated by simulating explicitly every physical state through which the system goes, or by reproducing the outcomes with shortcuts without following all the states of the system. However, in computationally irreducible processes, there is no shortcut that may provide an explicit algorithm which would connect the outcome with the input and therefore would stand for a possible explanation of the outcome. In such cases, the behavior of the system can only be found by direct simulation or observation.

2. I borrow this notion from Dubucs (2006) who characterizes derivation in formal systems "as the process of unfolding the mathematical content of the axioms by means of the progressive application of the inference rules." He adds that "running a computer program can be viewed as unfurling the content implicit in its instructions."

That said, we may think that, even if it is not possible for the user to go all over the simulation in *extenso*, it may still be possible for her to grasp some parts of the simulation and thereby to find the relevant explanatory components. This is the third and last account of how to gain relevance that I will consider now. Here searching for relevance requires to distinguish between relevant and irrelevant details in the computational processes. However, there is another kind of gap here due to another form of opacity that may also prevent the user from identifying the important elements through the simulation processes. This was later suggested by Humphreys:

> Here a process is epistemically opaque relative to a cognitive agent X at time t just in case X does not know at t all of the epistemically relevant elements of the process. A process is essentially epistemically opaque to X if and only if it is impossible, given the nature of X, for X to know all of the epistemically relevant elements of the process. (Humphreys 2009, p. 618)

I would like to make sense of this form of opacity (and to distinguish it from the first one) by introducing the notion of "explanatory noise."[3] If the user tries to connect the simulation outputs with the model components by going through the simulation processes, she would have to consider details which are relevant for explanations as well as details which may be important for computational purposes but remain useless for explanations. In other words, she would have to encounter explanatory noise that I define as being composed of all details that are necessary for the accuracy of calculations, but are irrelevant in the search for explanation by a cognitive agent. For instance, algorithms involve conditional loops—e.g., *if* or *while*—so that at each step the user would have to take into account all the required conditions and therefore would hardly grasp what is going on in the simulation. The situation is worse when simulations are used to study complex systems (that are described with many variables or are governed by non-analytical equations). Because, here, the user has to take into account an overwhelming number of data in order to follow the details of calculations.

The more one encounters explanatory noise, the more difficult it is for a cognitively unaided human to grasp relations of explanatory relevance between the inputs and the outputs of simulations. For that matter, this is the reason why Humphreys (2004) suggests that the epistemic opacity of simulations (in the first sense) is not necessarily a defect insofar as the

3. The term "explanatory noise" can be found in Batterman (2002) with a different meaning though.

negligence of details improves the understanding of the simulated systems. It seems to be the case that more detailed simulations always have a higher level of epistemic opacity. Detail and possibility in explanations seem to be inversely correlated.

The gap between the simulation outputs and the underlying model is a philosophical concept that nonetheless captures a real difficulty for scientists. This is well illustrated by the following testimony from oceanographer Achim Wirth:

> As simulation models are mere products of the human creation, it seems clear that our understanding of nature is almost perfect because we can calculate or imitate almost perfectly this nature. Nevertheless the expression "I understand what I have created" does not represent reality. The creation is done from understood components as computational models are built from equations expressing the fundamental laws of physics that we understand. *But the gap between these physical laws and the results of an integrated realistic global oceanic model is often too big to allow for a genuine human understanding*. The definition of "human understanding" by a scientist that I adopt is the following one: we have understood a process if we can explain with words its functioning as well as its reaction to variations of initial conditions or external parameters. (Wirth 2010, p. 5; emphasis added)

In a nutshell, the gap between the simulation outputs and the underlying model may be due to the lack of analytic understanding. It may be due to two sources of opacity: speed of simulations and absence of short-cuts. Or it may be due to explanatory noise.

Gap between simulation outputs and the underlying model is an obstacle for one to identify relevant explanatory components within a model. Therefore, it prevents the user getting explanations with simulations based on the model components and computational details. However, in the remainder of this paper, I will contend that the search for relevant explanatory components may be nevertheless (at least partly) possible via visual representations (e.g., graphs, pictures, films) that display the simulation outputs. I will then suggest that this is due to the fact that, when appropriately built, visual representations can saliently exhibit, with vivid colors for example, the relevant pieces of information and ignore the others.

3. Visual Representations of Simulation Outputs

Visual representations are often used in computer-assisted sciences. For example, they are ubiquitous in aeronautics, e.g., for describing the trajectory of spaceships; in molecular chemistry, e.g., for studying the geometrical

structures and the chemical composition of proteins or DNA, and also in meteorology and geography.

Their use in computer-assisted sciences is made possible by the development of computer graphics and visualization techniques. The creation of recent journals that deal specifically with these techniques, e.g., *Computer Vision and Image Understanding* by Elsevier and *Computing and Visualization in Science* by Springer, shows how important visual representations are in this scientific field.

For computer graphics researcher Thomas Defanti and computer scientist Maxine Brown, "access to visualization is critical" because "it does not simply represent the best way to look at data; it represents the only way to see what is going on" (DeFanti and Brown 1991, p. 258). Because computer simulations yield a great amount of data, the spatialization of these data under, e.g., the form of tables, graphs, or pictures, make the interpretation and the analysis of the data possible. Using an exclusively numerical format, the human brain could not interpret gigabytes of data each day. Thus, for them,

> A technical reality today and a cognitive imperative tomorrow is the use of images. The ability of scientists to visualize complex computations and simulations is absolutely essential to ensure the integrity of analyses, to provoke insights, and to communicate those insights with others. (Defanti and Brown, 1991, p. 252)

It therefore appears that visual representations play a crucial role in computer-assisted sciences in that they help us to interpret and to process the data that computer simulations provide. Some philosophers have also recognized that visual representations may be actually used to enhance understanding through simulation models (Humphreys 2004, pp. 110–14; Lenhard 2006; Carusi 2011, 2012; Boumans 2012).

In what follows, I will argue that, by making the many simulation data cognitively accessible, visual representations allow one to answer why-questions about the phenomena being studied. In order to do so, I will develop a case study. I will elaborate on the way how graphs and pictures can help one to explain the behavior of a flow past a solid cylinder from computer simulations. In particular, I will focus on explaining why vortices are induced by the interaction between the cylinder and the fluid at a certain threshold of the fluid inflow velocity.

The phenomenon of vortices in a flow past an obstacle has received and still receives great attention from scientists and engineers in view of the fact that the periodic onset of vortices can sometimes lead to unwanted structural vibrations which cause fatigue damage of the obstacle. These vibrations, called vortex-induced vibrations, can have serious consequences

when, for instance, they occur against bridges, offshore structures, or marine cables (see Sarpkaya 2004 for a review on theoretical, experimental, and numerical progress made the past two decades on this issue).

For the sake of simplicity, let us consider the very basic study of the flow past a cylinder described in terms of a bi-dimensional uniform flow of velocity vector v in a laminar regime (therefore non-turbulent) without heat transfer. The fluid is considered to be incompressible and slightly-viscous (e.g., air and water are slightly-viscous fluids). It circulates around a solid, smooth, and fixed cylinder of diameter D whose axis is perpendicular to the direction of the flow.

Scientists studied the behavior of a flow past an obstacle prior to the availability of computer simulations (see, e.g., the work of Adhémar Barré de Saint-Venant or Ludwig Prandtl in Darrigol 2005). But computer simulations provide great benefits here. First of all, without computer assistance, this study requires a heavy experimental setup, e.g., a water (or wind) tunnel coupled with methods of flow visualization such as the particle image velocimetry technique. Furthermore, simulations generate as many configurations of the system as desired, like for example configurations of the system at different length scales (e.g., width and length of the tube, dimension of the cylinder), for various geometrical configurations (e.g., circular, squared, or diamond-shaped cylinder), for different parameters (e.g., fluid viscosity, fluid density, Reynolds number). Therefore, it comes as no surprise that students in physics or in engineering are commonly introduced to this elementary study directly by a numerical approach.

Nevertheless, despite the benefits of simulations here, there is a gap between the simulation model and the simulation outputs that makes difficult for the user to get access to relevant explanatory components. In order to illustrate the gap, let me introduce the model in question under its computational form, i.e. ready for numerical computation. The model must contain the equations that govern the behavior of the fluid, and notably the equation of constraints $\vec{div}(\bar{\bar{\sigma}}) + \vec{f} = \vec{0}$.

The stress tensor defined by $\bar{\bar{\sigma}}(P, \vec{\nabla} v) = -P.Id + \mu.Tr(\bar{\bar{D}}).Id + 2.\eta.\bar{\bar{D}}$. P is the hydrostatic pressure, μ the dynamic viscosity, η the shear viscosity, and Id the identity matrix. The displacement gradient tensor is defined as:

$$\bar{\bar{D}} = \begin{bmatrix} \dfrac{\partial v_x}{\partial x} & \dfrac{1}{2} \cdot \left(\dfrac{\partial v_x}{\partial y} + \dfrac{\partial v_y}{\partial x} \right) \\ \dfrac{1}{2} \cdot \left(\dfrac{\partial v_x}{\partial y} + \dfrac{\partial v_y}{\partial x} \right) & \dfrac{\partial v_y}{\partial y} \end{bmatrix}$$

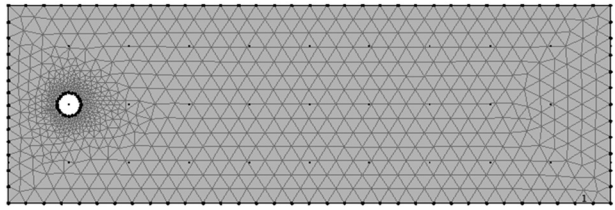

Figure 1. Example of meshing

The components of the force exerted on the fluid by the cylinder are:

$$f_x = -\rho\gamma_x = -\rho\frac{\partial v_x}{\partial t} = -\rho\left(\frac{\partial v_x}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial v_x}{\partial y}\frac{\partial y}{\partial t}\right) = -\rho\left(v_x\frac{\partial v_x}{\partial x} + v_y\frac{\partial v_x}{\partial y}\right)$$

$$f_y = 0$$

The boundary condition near the cylinder is the no slip condition ($\vec{v} = \vec{0}$). And the fluid is considered as incompressible (consequently $di\vec{v}(\vec{v}) = \vec{0}$).

In order to solve the equations, one must work out a numerical scheme based on the finite element integration of the equations. First the finite element method consists in discretizing the study domain into finite elements. Figure 1 displays an example of two-dimensional meshing of the domain with triangular elements.[4] We can notice that the finite elements are smaller near the cylinder wall in order to optimally take into account the effect of viscosity on the fluid's behavior (e.g., boundary layer phenomenon). Since the finite elements are triangular, each element is composed of three nodes. Each node j is associated with a function β_j generated by the concatenation of the Lagrange polynomials that are defined for every finite element whose one apex is the node j. This function equals one at the node j, and zero at the other nodes. The total number of nodes in the domain is usually in the order of several thousand or millions of nodes depending on the meshing.

4. Here are the selected geometrical dimensions that I used for the simulations: the cylinder diameter is 4 mm; the domain width is 3.4 cm; and the domain length is 10 cm. For the design of the geometry, the meshing, and the simulations I have used the software COMSOL Multiphysics™.

Under the weak formulation, the system of equations to solve become:

$$(2\eta\frac{\partial v_x}{\partial x}-p)\frac{\partial \beta_j^x}{\partial x}+\eta(\frac{\partial v_x}{\partial y}+\frac{\partial v_y}{\partial x})\frac{\partial \beta_j^x}{\partial y}+\rho(v_x\frac{\partial v_x}{\partial x}+v_y\frac{\partial v_x}{\partial y})\beta_j^x=0$$

$$\eta(\frac{\partial v_y}{\partial x}+\frac{\partial v_x}{\partial y})\frac{\partial \beta_j^y}{\partial x}+(2\eta\frac{\partial v_y}{\partial y}-p)\frac{\partial \beta_j^y}{\partial y}+\rho(v_x\frac{\partial v_y}{\partial x}+v_y\frac{\partial v_y}{\partial y})\beta_j^y=0$$

$$\frac{\partial v_x}{\partial x}+\frac{\partial v_y}{\partial y}=0$$

v_x, v_y, and p are the three variables of the system; they are the two components of the fluid velocity and its pressure. The newly obtained equations are then integrated at each element, i.e. at each βj, and at each time step when a time-dependent module is needed. For that purpose, for each node j, the partial differentials of the variables are replaced by approximations which are calculated with the values of the variables on the nodes adjacent to the node j.

Let me now develop on why there is a gap between the simulation outputs and the model in this particular case. First the weak formulation hardly gives an analytic understanding in that the terms are complex so that one cannot easily figure out how the components interact with each other during the simulations. Second the simulation runs fast on the computer. It only lasts a few minutes, thus making it impossible for a human being to follow the computational processes. This is what I have identified as a source of opacity. Furthermore, when the simulation runs, the number of calculations is proportional to the number of nodes and the time steps, as well as the number of equations to solve. Thus, even if it was possible to slow down the simulation it would take too long to survey every calculation step. Third, there is a gap due to explanatory noise in that many computational steps, e.g., the calculations at each node, are not relevant for explanations.

Because of the gap, here, it seems to be hardly possible to identify within the model the relevant explanatory components that may explain the behavior of the flow and especially the onset of vortices. That said, as I will now show, visual representations help here. The example illustrates how in practice the user explains simulated phenomena via visual representations.

The starting point of the user's investigation consists in selecting the variables she wants to edit from those she deliberately disregards. In other

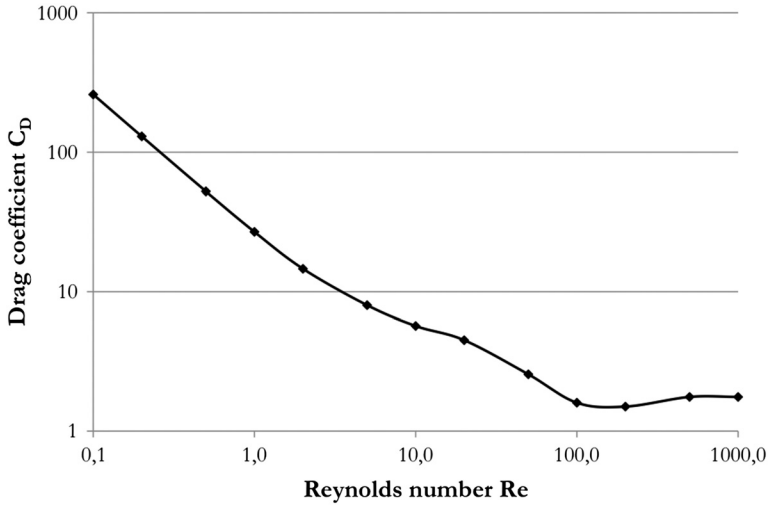

Figure 2. Variation of the drag coefficient against the Reynolds number

words, before running a simulation, the user should already know the variables of interest. This is an important stage since visual representations are then used to represent the relations between those variables. In the example, a user should identify the physical magnitudes which are involved in flow instabilities. She should be interested in the physical implications of the cylinder on the fluid, and therefore in the evolution of the force exerted by the surface of the cylinder on the fluid against the variation of the mean inflow velocity, or more commonly against Reynolds number.[5] The two components of this force are the drag, which is parallel to the oncoming flow direction, and the lift, which is perpendicular to it.[6] Further, she should also be interested in the fluid velocity varying against Reynolds number. Consequently, she would notably select the drag and lift coefficients, Reynolds number, and the fluid velocity as the key variables she wants to edit.

Afterwards, when the simulation ends, the user generally extracts these outputs from the outputs files by hand or by using a specific subprogram: this is called post-processing. Then she displays the extracted outputs, or

5. Reynolds number is a dimensionless parameter defined as $Re = (\rho.v.D_H) / \eta$ with ρ the fluid density (kg/m^3), v the mean inflow velocity (m/s), D_H the hydraulic diameter (m), and η the dynamic viscosity of the fluid (kg/(m.s)). The properties of water are the following ones: mass density $\rho = 10^3$ $kg.m^{-3}$ and dynamic viscosity $\eta = 10^{-3}$ Pl.

6. The drag coefficient C_D is defined as $C_D = |F_x| / ((1/2).\ \rho.\ v_0^2.\ D_H)$ and the lift coefficient C_L is defined as $C_L = |F_y| / ((1/2).\ \rho.\ v_0^2.\ D_H)$.

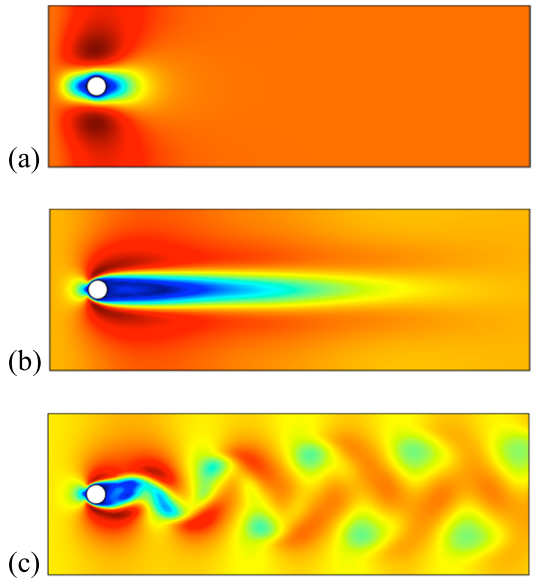

Figure 3. Flow regimes—(a) Re < 5, (b) 5 < Re < 50 and (c) 50 < Re < 300

their averages, through visual representations, i.e., graphs, pictures, or diagrams. For example, from the simulation outputs, she can extract the drag coefficient values, and then can draw in a graph of the evolution of the drag coefficient against Reynolds number. From the graph $C_D(Re)$, in figure 2, it results that $C_D(Re)$ is a decreasing function. Consequently, she can draw an initial piece of information: the higher Reynolds number is, the less the drag has influence on the fluid. Thus the user learns that an increase in fluid inflow velocity goes with an increase in the cylinder resistance against the fluid, and consequently, provokes a distortion of its own trajectory, enabling flow instabilities.

For each relevant range of Reynolds number the user can also plot the fluid velocity fields projected onto the system geometry. Once created, the three velocity fields in figure 3 give other pieces of information. These fields are representative of the three following distinctive ranges of Reynolds number: $0 < Re < 5$; $5 < Re < 50$; $50 < Re < 300$.[7] The color scale, from blue to red, represents the variation of the fluid velocity values, from the lowest to the highest.

7. These approximate values of ranges of Reynolds number have been assessed with my simulations. They must not be considered as reliable values although they are not aberrant.

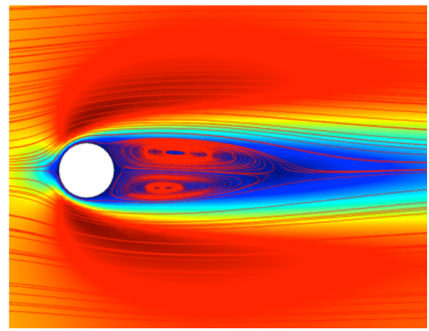

Figure 4. Streamlines ($5 < Re < 50$)

From the three pictures in figure 3, we can gain new information about the three specific laminar flow regimes. Figure 3.a—for $Re < 5$—displays a creeping flow for which the velocity field profile is symmetrical and for which there is no boundary layer separation, i.e., no detachment of the fluid portion closest to the cylinder wall in flow direction. In figure 3.b—for $5 < Re < 50$—two fixed contra-rotative vortices are formed in the wake of the cylinder. To see them, we need to plot the streamlines as they are in red in figure 4.

Other simulations, with increasing Re, that are not presented here, show that as the vortex formation lengthens, the symmetry of the fluid flow spontaneously breaks down, leading to instabilities. Eventually there is a separation of the boundary layer, generating a discrete vortex. At this point a vortex-shedding phenomenon occurs in which vortices are shed alternatively and periodically at either side of the cylinder. Figure 3.c—for $50 < Re < 300$—shows the induced regular pattern of the double row of vortices in the wake of the cylinder; it is called the von Kármán vortex street. The frequency of this vortex shedding depends on Reynolds number.

Finally, from figure 3 and figure 4, we can obtain important pieces of information about the behavior of the flow past a cylinder. In particular they are helpful in answering why a vortex is emitted when Reynolds number is superior to 50. The following explanation can be given, which is inspired from Sumer and Fredsøe (2006). First the boundary layer separates from the cylinder surface because the divergent geometry of the flow environment at the rear side of the cylinder imposes an adverse pressure gradient—as shown in figure 5. Because of this separation, a shear layer is formed on the right of the separation point (see figure 7). Lastly, the high vorticity, i.e., circulation per fluid surface, initially contained in the boundary layer feeds the shear layer as shown in figure 6. Consequently, the shear layer rolls up into a discrete vortex.

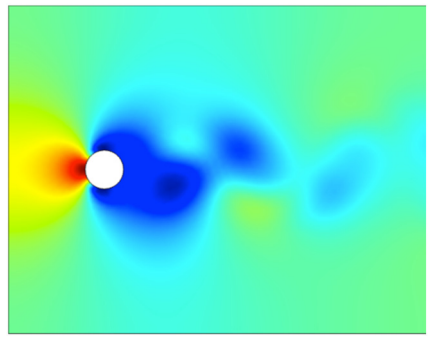

Figure 5. Pressure gradient

Each time, a visual representation provides pieces of information which are used to answer a specific why-question. Depending on the question, new appropriate visual representations may be built which aim at providing relevant pieces of information. Nevertheless we should not think that visual representations are only useful in a computer-assisted context. Visualization was already important for Ludwig Prandtl in applying the Navier-Stokes equations to the flow past a cylinder. It is striking how similar the scheme of velocity streamlines in the work of Ludwig Prandtl is to the scheme in figure 4 (Darrigol 2005; Prandtl 1927). Visualization has been helpful to Prandtl for creating his model, allowing him in this way to provide approximate solutions to Navier-Stokes equations for inviscid incompressible steady flow, and to introduce and explain the boundary layer phenomenon (Heidelberger 2006). It even seems that Prandtl developed a certain intuition, a "visual understanding" of the phenomena

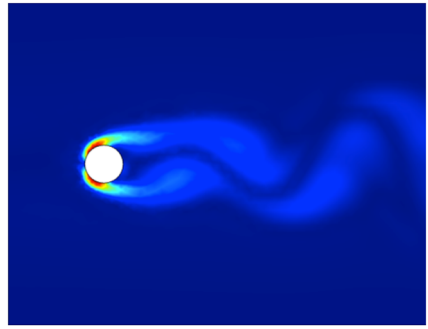

Figure 6. Vorticity magnitude

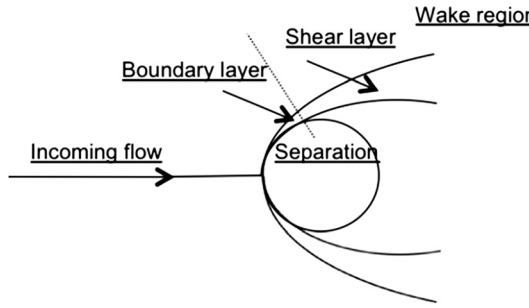

Figure 7. Flow regions (diagram built from Figure 1.2 and Figure 1.5 of Sumer and Fredsøe 2006)

before even setting the equations of his model. On that subject, Prandtl said:

> Herr Heisenberg has [...] alleged that I had the ability to see without calculation what solutions the equations have. In reality I do not have this ability, but I strive to form the most penetrating intuition [*Anschauung*] I can of the things that make the basis of the problem, and I try to understand the processes. The equations come later, when I think I have understood the matter. (quoted in Darrigol 2005, p. 287)

In the example, explanations of the flow behaviors under different regimes from simulations are found via investigation of visual representations. The representations bring pieces of information which derive from simulations but are not available from the model or the computational processes. These pieces of information are explanatory components about the main characteristic behaviors of the simulated phenomena which are in turn used in answering why-questions. For example, they allow one to reconstruct the mechanism of the onset of the first vortex. Yet being able to get any explanation about the simulated phenomena is another matter. Visual representations do not completely fill the gap but partly fill the gap in representing relations between relevant variables which vary all along the simulation process and whose evolution cannot be tracked through the computational processes. Visual representations succeed in conveying relevant explanatory components in such a context because they have some specific properties that I will now discuss.

4. What makes Visual Representations Important

In this section, I will discuss the properties of visual representations that make them suitable for explanatory tasks in a computer-assisted context.

Visual representations are of course useful here partly due to the fact that our visual device is well adapted for analyzing pictures and in particular for recognizing visual patterns. They also have three intrinsic properties that make them particularly adequate for exploring the relevant explanatory components within a simulation model. First visual representations can be made of a great amount of data, thus, being synoptic views of these data, they can allow the user to get access to several pieces of information about the simulated phenomena at a single glance. Second, they can reveal the minimal useful amount of information contained in computer simulations that the user needs to know; they do so by exhibiting the relationships between the appropriate variables. Third, they make this information easily cognitively graspable in that they represent it in a salient manner.

4.1. Relations of Dependence

Before discussing these three properties, I want to suggest that visual representations, built from simulation outputs, are generally presentations of relations of dependence between variables of interest. This corresponds to the practice of post-processing simulations which often consists in drawing on visual representations the variations of relevant variables against other relevant variables. In this way, visual representations allow the user to identify relations of dependence. This definition is supposed to work for, at least, tables, graphs, pictures or films used in a simulation context.

In a table made of numbers, numerical data are arranged in the checkered visual support (e.g., paper or computer screen) so that they are in relation to each other. Each data is placed at a specific position within the table and this position is also associated with another variable. For example, let us consider table 1, a table made of the values of drag coefficient C_D against Reynolds number. C_D is made in relation with Re because each datum of C_D aligns a datum of Re.

In a graph, a picture, or a map, the data are also arranged in the visual support. But instead of being under an alphanumeric form, they are represented by what I shall call an "iconic mark," e.g., point, dash, dotted line, asterisk or colored area. These iconic marks are substitutes for some particular numerical datum and are located in the space of the visual support. Thus they stand for the datum itself and also for the relation of

Table 1. Drag coefficient C_D against Reynolds number Re

Re	0,1	0,2	0,5	1	2	5	10	20	50	100
$\mathbf{C_D}$	260	130	52	27	15	8	6	5	3	2

dependence between this datum and the variable(s) that the spatial coordinate of the visual support (paper or computer screen) represents. For example, the mark "\" in the graph (figure 2) stands for the relation between C_D and Re. Colored areas (blue, yellow, orange or red) in the flow pictures (figure 3) stand for the velocity against the actual spatial coordinates of the flow. A simulation film can also be seen as a presentation of relations of dependence in which the time in the film represents the time in the phenomenon under study.

My suggestion, i.e., to define visual representations as presentations of relations of dependence, seems to match the way scientists also conceive visual representations. In many cases, as Defanti and Brown highlight considering in particular simulation films, "Scientists want to compute phenomena over time, create a series of images that illustrate the interrelationships of various parameters at specific time periods" (Defanti and Brown 1991, p. 253). Now that I have made the suggestion that visual representations are presentations of relations of dependence, I will now discuss the three properties that make them valuable means of finding relevant explanatory components.

4.2. Synoptic Views

Making accessible a great amount of data at a first glance is a first general property of visual representations. It is not specific to their use in simulation context (Tufte [1990] 2001). In order to illustrate this property, statistician Tufte (2001) presents a picture of the northern galactic hemisphere (figure 8

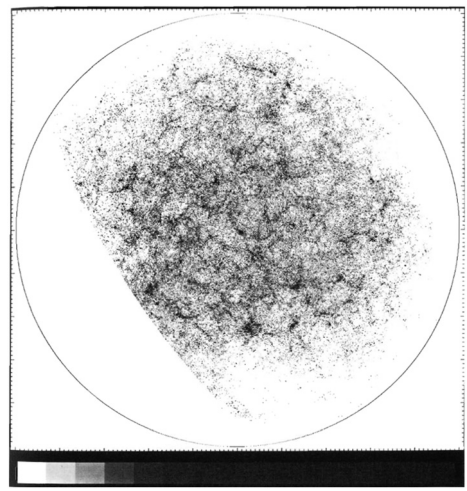

Figure 8. The northern galactic hemisphere (Seldner et al. 1977, reproduced in Tufte 2001, p. 27)

on the left).[8] The picture displays the distribution of 1.3 million galaxies. Tufte specifies that it "divides the sky into 1,024 × 2,222 rectangles. The number of galaxies counted in each of the 2,275,328 rectangles is represented by ten gray tones; the darker the tone, the greater the number of galaxies counted" (Tufte 2001, p. 26). Tufte writes about the map of the norther galactic hemisphere that "the most extensive data maps [...] place millions of bits of information on a single page before our eyes. No other method for the display of statistical information is so powerful" (Tufte 2001, p. 26).

In the pictures of flow (figure 3), as many data are conveyed as there are nodes in the computational model. More than that: some data which are interpolated from calculated data are represented there. Indeed colors have been assigned to the points in the visual support for which there was no performed calculation; these points are those that are not placed on the nodes of the grid. The same is true for graphs. In the graph, the continuous curves go through the calculated data as well as through interpolated ones.

The most interesting aspect here is that a visual representation can exhibit a great amount of data in a structured way. Unlike a messy mass of data, from which it would be hard to draw relevant pieces of information, appropriate visual representations enable one to infer specific pieces of information about the target system. This is due not only to the fact that they can convey a great amount of data but also to the fact that they are, as I have suggested, presentations of relations of dependence. They arrange within the space of the visual support the many data so that these data can be in relation with each other. In figure 8, galaxies are arranged within the space of the northern galactic hemisphere so that the figure gives the relation between the existence of galaxies and the space.

This is an important property for the purpose of finding explanations since explaining with a simulation involves recognizing the relevant pieces of information among the simulation outputs. Let me illustrate this by considering a user who wants to determine where the lowest fluid velocities in a flow are. Without visual representations, she would have to examine the vast amount of simulation outputs, and detect from this the important pieces of information, i.e. the nodes in which the velocities are the lowest, and then find out to which positions these nodes correspond to in the system. This task would of course be tedious if not impossible. By contrast, visual representations would enable the user to find the positions

8. The north galactic pole is at the center of the picture. The sharp edge on the left results from the earth blocking the view from the observatory. Around the picture, the view is obscured by the interstellar dust of the Milky Way.

immediately in virtue of displaying a great amount of data in a synoptic manner.

This property is specific to visual representations. Kulvicki (2010) compares the way graphs and images on the one side, and linguistic descriptions on the other, deliver information. From this comparison he concludes that:

> Graphs and images [...] are good at providing fine-grained, quantitative information [...]. They can... present vast amounts of information about a great many features of their objects. The plausible story is that images and the like deliver a lot of rather specific information while descriptions and their ken are able to deliver arbitrarily little. If we only need a little bit of information, descriptions are the superior means of conveying it, but if we have a lot of very specific information and we want to deliver it, images are best. (Kulvicki 2010, p. 297)

Consequently, in computer-assisted sciences, where the amount of computed data is overwhelming, graphs and images are justifiably very useful tools for presenting data. In conveying a great number of data, they allow us to draw pieces of information that one would hardly detect from the numerical data that simulations generate.

That said, there is a limit, of course, in using visual representations. As Tufte highlights, "not a great many substantive problems [...] are exclusively two-dimensional. Indeed, the world is generally multivariate" (Tufte 1997, p. 17). Thus we can visually present a three-dimensional object on a picture only by projecting it into a plan. Therefore, the concrete frame of a visual representation, as well as our visual device, restrict the number of variables which can be represented in the visual representation.

4.3. Revealing Power

Visual representations not only have the property to convey a great amount of data, but they can also reveal important pieces of information that we can draw from these data. Revealing power is the second property I want to emphasize. On the one hand, visualization enables one to recognize the effects of the postulated theoretical structures that are constitutive of the underlying theory. In the example, visualization enables one to recognize the effects of the postulated theoretical structures that are constitutive of hydrodynamic theory such as the boundary layer, the fluid separation, the shear layer, or the wake. Without visual representations, we would have the greatest difficulty in distinguishing the flow regions in the mass of

numerical data generated by simulation. We would not be able to even assess at which Reynolds number the first vortex is produced since we would not see any vortex. On the other hand, visual representations reveal the pieces of information contained in simulations by making them extractable. In other words, information stands out on visual representations, enabling us to cognitively master it.

That said, for extractability to be met, visual representation must satisfy a condition outlined by Kulvicki (2010). (We should note that Kulvicki studies graphs and images in general, not only the ones used in a simulation context). The pieces of information in visual representations are extractable if each syntactic feature of visual representations—e.g., color, shade, shape—aims to convey a unique piece of information, no more. For example, figure 3 presents extractable pieces of information in that, for the colored regions, "being blue," "being green," or "being red" are the features of the pictures which exclusively indicate specific ranges of fluid velocities. They carry no other information, for example, about the localization and the density of the areas of each range of velocities. Other features of the pictures, such as the relative position and the surface of the colored regions, are responsible for that.

I want to suggest that extractability is also made possible by the use of iconic marks. All visual representations (except tables) have iconic marks. These marks are useful in extraction of information because they carry not only a datum but also an additional piece of information that is not explicitly contained in the data but corresponds to a product of an inference from these data. For example, the mark "\" in the graph (figure 2) not only represents data but also symbolizes a decrease of the drag coefficient C_D against Reynolds number Re; while the mark "/" would symbolize an increase. These marks thus deliver directly an immediate piece of information that would be less easy—although not impossible—to infer from the mass of data. Other examples are the colored areas in the flow pictures. These marks indicate immediately where, for example, the area of lowest velocity behind the cylinder is (this is the blue area), or where the onset of a vortex is.

Consequently, it seems that some iconic marks can themselves convey more than the data since they are perceived features which indicate immediately the results of (longer) inferences that one would have drawn in browsing the table of data. This immediacy of the extraction of these results is allowed by our own ability to recognize iconic marks, but also by their irreducibly conventional (and thereby cultural) nature. For these markers to be interpreted, the designer of the representation must specify the rules of interpretation (e.g. legends, scales). But sometimes this is not needed since some typical iconic marks are well-known and their

interpretation has become virtually automatic habit. It even seems that this interpretation needs no effort. As Goodman writes, "practice has rendered the symbols so transparent that we are not aware of any effort, of any alternatives, or of making any interpretation at all" ([1968] 1976, p. 36).

4.4. Selective Function

Visual representations make pieces of information readily available. A third property of visual representations that I will now discuss is that, under certain conditions, they can have a selective function in that they can make only the relevant pieces of information readily available, and ignore the others. This function, as I will show, helps precisely to overcome the problem of explanatory noise.

Recall what is at stake. In order to have access to the evolution of relevant variables, one needs to run simulations and to plot these variables of interest against others. But, if one wants to identify their evolution in the detail of numerical calculations, one has to consider relevant details as well as irrelevant details, and therefore encounters explanatory noise. However visual representations can overcome the problem of explanatory noise if they are selective. They can saliently exhibit, with vivid colors for example, the relevant pieces of information which are necessary for the purpose of reconstructing explanations about the behavior of the system under study, and ignore the others irrelevant details. They are selective, though, if they satisfy two conditions which are semantical salience and syntactical salience (Kulvicki 2010).[9]

Kulvicki (2010) defines the condition of semantic salience as follows: syntactic features of visual representations and pieces of information must be well correlated. In order for a part of a representation to be semantically salient, it must be straightforward for a user to determine what that part represents: "there must be a plan of correlation between features of the representation and features of the data that is easy to grasp" (Kulvicki 2010, p. 301). Depending on what kind of explanation one wants to draw from visual representations, one needs to select the proper relationship between perceptual features of visual representations and the kinds of information to display. If, for example, one wants to know the regime of a flow, one would want to see the global and spatial picture of fluid velocities, like in figure 3,

9. These two concepts are defined by Kulvicki (2010) as conditions for immediacy (the property of any representation to make some aspects of its content immediately available). He also adds the extractability of information as a condition for immediacy. But here I want to suggest that semantical salience and syntactical salience are also conditions for the visual representations to select relevant pieces of information among other details.

and must ensure that each range of fluid velocities is well represented by a distinct color.

The choice of the colors, and more generally the choice of syntactic features, must also be judicious. Kulvicki calls this condition the search for syntactic salience. The choice of syntactic features directly affects our ability to extract information from representations. If, for instance, the highest range of velocities is in vivid red, the lowest in crimson, and the intermediate ranges in shades of red, it would be more delicate for us to distinguish between the distinct colored regions on the visual and to comprehend their variation than, for example, the more contrasting colors of the rainbow. Consequently, we will have more difficulty in figuring out which regions of the system under study possess a certain range of velocity.

Therefore figure 3 is a good example of visual representations that meets these two requirements. The pictures only represent, with vivid colors, the gradient of fluid velocity field against space and for each different range of Reynolds number. The selection of the relevant variables is prepared by the user, but then visual representations can only make it possible to present those variables. For that purpose, they do not present all the approximations of the parameters and of the variables (e.g., pressure, forces) needed for the calculation of each velocity in every finite element. Thus, the explanatory noise constituted by the calculations of those parameters is "filtered" by the visual representations. Only the pieces of information, which are deemed relevant in explanations about the simulated system, are perceptually highlighted by salient syntactic features.

In sum, if they meet syntactic salience and semantic salience, visual representations can make readily available the pieces of information from computer simulations which are relevant for the reconstruction of explanations about the system under study. Thus, while there is a gap between the simulation model and simulation outputs, visual representations still enable one to obtain explanations.

5. Conclusion

I have shown that there is a gap between the simulation outputs and the underlying model that prevents users from getting explanations. I have also argued that visual representations can nevertheless be used to access the relevant explanatory components. I have illustrated this possibility via the example of the study of a flow past an obstacle. Then I have specified the reasons that make visual representations particularly suitable for explanatory tasks in the computer-assisted sciences: they can represent the great amount of data that computer simulations generate, they reveal the pieces of information that are hidden in the opaque simulation process,

and, under certain conditions, they can make readily available only the pieces of information that matter for the reconstitution of explanations.

Visual representations are therefore precious means of conveying simulation outputs. That said, they probably do not completely fill the gap between the simulation outputs and the underlying model. There may be other inconvenience due to opacity and explanatory noise that the use of visual representations does not avoid.

References

Batterman, Robert W. 2002. *The Devil in the Details, Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

Bokulich, Alisa. 2009. "Explanatory Fictions." Pp. 91–109 in *Fictions in Science: Philosophical Essays on Modeling and Idealization*. Edited by M. Suárez. London: Routledge.

Bokulich, Alisa. 2011. "How Scientific Models Can Explain." *Synthese* 180 (1): 33–45.

Boumans, Marcel. 2012. "Visualisations for Understanding Complex Economic Systems." *Ways of Thinking, Ways of Seeing. Automation, Collaboration, & E-Services* 1: 145–165.

Carusi, Annamaria. 2011. "Computational Biology and the Limits of Shared Vision." *Perspectives on Science* 9 (3): 300–336.

Carusi, Annamaria. 2012. "Making the Visual Visible in Philosophy of Science." *Spontaneous Generations: A Journal for the History and Philosophy of Science* 6 (1): 106–114.

Clark, Andy. 1987. "The Kludge in the Machine." *Mind and Language* 2 (4): 277–300.

Darrigol, Olivier. 2005. *Worlds of Flow: A History of Hydrodynamics from the Bernoullis to Prandtl*. Oxford: Oxford University Press.

DeFanti, Thomas A., and Maxine D. Brown. 1991. "Visualization in Scientific Computing." *Advances in Computers* 33: 247–307.

De Regt, Henk, Sabina Leonelli, and Kai Eigner (eds.). 2009. *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

Dubucs, Jacques. 2006. "Unfolding Cognitive Capacities." Pp. 95–101 in *Reasoning and Cognition*. Edited by M. Okada. Keio: Keio University Press.

Frisch, Mathias. 2015. "Predictivism and Old Evidence: a Critical Look at Climate Model Tuning." *European Journal for Philosophy of Science* 5 (2): 171–190.

Goodman, Nelson. [1968] 1976. *Languages of Art. An Approach to a Theory of Symbols* 2nd edn. Indianapolis: The Bobbs-Merrill Company, Inc.

Heidelberger, Michael. 2006. "Applying Models in Fluid Dynamics." *International Studies in the Philosophy of Science* 20 (1): 49–67.

Humphreys, Paul. 2004. *Extending Ourselves*. Computational Science, Empiricism, and Scientific Method. Oxford: Oxford University Press.
Humphreys, Paul. 2009. "The Philosophical Novelty of Computer Simulation Methods." *Synthese* 169 (3): 615–626.
Kulvicki, John. 2010. "Knowing with Images: Medium and Message." *Philosophy of Science* 77 (2): 295–313.
Kuorikoski, Jaakko. 2011. "Simulation and the Sense of Understanding." Pp. 250–273 in *Models, Simulations, and Representations*. Edited by Paul Humphreys and Cyrille Imbert. London: Routledge.
Lehtinen, Aki, and Jaakko Kuorikoski. 2007. "Computing the Perfect Model: Why Do Economists Shun Simulation?" *Philosophy of Science* (74): 304–329.
Lenhard, Johannes. 2006. "Surprised by a Nanowire: Simulation, Control, and Understanding." *Philosophy of Science* 2006: 605–616.
Lenhard, Johannes, and Eric Winsberg. 2010. "Holism, Entrenchment, and the Future of Climate Model Pluralism." *Studies in History and Philosophy of Science* Part B, 41 (3): 253–262.
McEvoy, Mark. 2008. "The Epistemological Status of Computer-Assisted Proofs." *Philosophia Mathematica* 16 (3): 374–387.
Morrison, Margaret. 2009. "Understanding in Physics and Biology: From the Abstract to the Concrete." Pp. 123–145 in *Scientific Understanding: Philosophical Perspectives*. Edited by H. de Regt, S. Leonelli, and K. Eigner. Pittsburgh: University of Pittsburgh Press.
Prandtl, Ludwig. 1927. "Über Flussigkeitsbewegung bei sehr kleiner Reibung." (Motion of Fluids with Very Little Viscosity), from Vier Abhandlungen zur Hydrodynamik und Aerodynamik. *National Advisory Committee for Aeronautics Technical Memorandum* (452): 1–8.
Salmon, Wesley C. 1998. *Causality and Explanation*. Oxford: Oxford University Press.
Sarpkaya, T. 2004. "A Critical Review of the Intrinsic Nature of Vortex-Induced Vibrations." *Journal of Fluids and Structures* 19: 389–447.
Seldner, Michael, B.H. Seibers, Edward Groth, and James E. Peebles. 1977. "New Reduction of the Link Catalog of Galaxies." *Astronomical Journal* 82: 249–314.
Sumer, B. Mutltu, and Jørgen Fredsøe. 2006. *Hydrodynamics around Cylindrical Structures*. Advanced Series on Ocean Engineering, vol 26. Singapore: World Scientific Publishing Co.
Tufte, Edward. 1990. *Envisioning Information*. Cheshire, Conn.: Graphics Press.
Tufte, Edward. 1997. *Visual Explanations*. Images and Quantities, Evidence and Narrative. Cheshire, Conn.: Graphics Press LLC.
Tufte, Edward. [1983] 2001. *The Visual Display of Quantitative Information*. Cheshire, Conn.: Graphics press.

Winsberg, Eric. 2012. "Values and Uncertainties in the Predictions of Global Climate Models." *Kennedy Institute of Ethics Journal* 22 (2): 111–137.

Wirth, Achim. 2010. *Etudes et évaluation de processus océaniques par des hiérarchies de modèles*. Research dissertation to obtain the "Habilitation à Diriger des Recherches," Laboratoire des Ecoulements Géophysiques et Industriels, Grenoble.

Wolfram, Stephen. 1985. "Undecidability and Intractability in Theoretical Physics." *Physical Review Letters* 54: 735–738.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.