# Describing Models

Before diving into Zollman's models, I want to explore models in general and to define terms that I will continue to use for the rest of this thesis. This section will touch on the vibrant debate about the role of models in science by explicitly presenting the interpretations I'm using to understand models. Overall, I do not endorse this set of definitions as the right or the only way to interpret models, I simply propose that this understanding is the right lens through which to understand the issues salient in computational modeling of the sort I discuss here. To this end, I will propose that I focus on representational mathematical models as a natural lens to view computational modeling. Next, I discuss the differing purposes of modeling and why I opt to focus on explanatory, rather than predictive modeling when discussing computational models. Finally, I will advocate viewing explanation in computational models as *mechanistic* explanation because computational models lend themselves well to elucidating mechanisms because they are highly decomposable.

## What Sort of Model Computes?

When I say a *model*, I refer to the mathematical structures used to represent some real phenomena in a target system. If we adopt the ontology for models proposed by Frigg and Hartmann, I refer to *representational* models of phenomena, or models which are positioned relative to a target. When a model M models a target T, we would say "M models T" and mean that M serves as an effective stand in for T in the context relevant to the goals of the model. It is important to distinguish these representational models from models which represent data points. A model of data can certainly inform a model of phenomena, but discussion of modeling data itself is typically discussed in the context of experimentation and empirical evidence, rather than the context of scientific understanding.[1]

In addition to being representational models of phenomena, I focus on mathematical models. By mathematical, I mean they are composed of equations rather than physical objects, set-theoretic structures, textual descriptions or any other kind of model. I focus on this sort of model because this sort of model lends itself best to direct computer simulation. Because computers at their core crunch numbers and move bits around, simulating a model on a computer requires converting that model into equations and mathematical structures that can be described as a computable program. Because the process of programming leads to simulated models being specified as equation-based models, the models-as-equations understanding is frequently used to discuss how computer simulations and their associated models work as a part of scientific inquiry.[2]

Furthermore, these relational models are inextricably tied to the possible worlds they model. A model has a *target*, which I broadly take to be an object or

[1] P. 11 Frigg and Hartmann, "Models in Science."
[2] Winsberg, "Computer Simulations in Science."

phenomena in a real or possible world that a model represents. Any equation-based representational model is going to somehow be about a target in some world, otherwise it would not be representational. To illustrate this, consider a scientist studying a tissue and its constituent cells. Say she has formed a model which represents the behavior of the cells within the tissue which accurately reflects the behavior of actual cells in the actual tissue. Such a model reflects the real world because there is an accurate representational relationship between the equations of the model and a tissue that exists in the real world. Now, say the scientist knows a drug affects cells in a certain way though some empirical means. She could form a model with the modified cell behavior and use it to explore how the tissue would behave under this regime. In this case, the model models the tissue in a hypothetical world in which the cells are affected by the drug.

Finally, I should clarify exactly what I mean by "simulation". I adopt Eric Winsberg's definition as "a program which runs on a computer and uses step-by-step methods to explore the approximate behavior of a mathematical model."[3] This definition stresses both that simulations approximate mathematical models and that they are typically time-dependent. The simulation approximates, rather than mimics perfectly the model because computers deal in discrete data so whenever a continuous model equation is modeled, some fidelity is lost. Consider modeling a flowing river. Where the model equations might define smooth rules for how the water flows at any time scale, a computer might need to simulate the model in second-long chunks of time for the computation to be tractable. If the system is already discrete, the simulation may more exactly approximate the model mathematics. Furthermore, this definition stipulates that simulations concern time-dependent processes, which rules out static computer renderings of geometrical structures, but allows videos of such renderings as could be the case in a virtual reality simulation. In practice, this focuses models on describing and understanding dynamic processes rather than static structure or other unchanging properties.

### Example: Modeling a Cannon

To further discuss how equation-based models represent phenomena in the real world, consider how one might formulate and use such an equation. Say an engineer, Alice, builds a cannon and places it on top of a fort. Alice wants to know how far the cannon can shoot to understand how far that cannon can fire a cannonball, so she might decide to use kinematic equations to model the cannonball's flight path. To do this, Anna wants to use as complete and as simple an account of motion as possible to understand what affects the flight path of the ball. Thus, she adapts Newtonian mechanics' kinematic equations to model the cannonball's motion. However, the a theory of mechanics on its own is insufficient to model the cannonball and Alice must chose how she will model the initial height and velocity of the ball, the angle of the cannon as well as the

[3] Winsberg.

acceleration due to gravity. Say Alice knows how tall the fort is (denoted $y_0$), the initial velocity of the ball for a standard shot ($v_0$), the acceleration due to gravity on Earth ($g$) and can set the cannon angle to $\theta$ degrees. If she assumes all these values are fixed, she can use the general kinematic equations to relate all these quantities. For example, the ball's distance from the ground could be modeled in relation to time as follows, where $v_0 \sin(\theta)$ is the vertical component of initial velocity:

$$y(t) = y_0 + v_0 \sin(\theta)t - g_0 t^2$$

Using this equation, Alice could solve for the time when the ball hits the ground, $t_h$. Alice can model this in the kinematic equation as some time after the ball is fired at $t = 0$ (meaning that $t_h > 0$) when the ball ends up on the ground (meaning $y(t_h) = 0$). Once this time is found, Alice can formulate another kinematic equation to model the horizontal motion of the ball, where $v_0 \cos(\theta)$ is the vertical component of initial velocity:

$$x(t) = v_0 \cos(\theta)t$$

Note that in the kinematic equation for horizontal movement does not include an initial position or a term for gravity because Alice models the distance away from the cannon over time and gravity is only modeled in the kinematic equation for vertical position as gravity principally accelerates objects towards the Earth. Now, using this equation-based model of the motion of cannonball, Alice can find that the ball travels a distance of $x(t_h)$.

What is important to note about this model is that Alice uses it as a stand-in for firing an actual cannonball in the actual world. Thus, Alice takes these mathematical equations to represent a real cannon shot despite them being relatively simplistic and ignoring many parts that might have some effect that is not deemed relevant for the behavior Alice wises to understand. For example, she doesn't model the affect of wind or the friction the ball would experience when the ball hits the ground. If we look closely, this means that the modeled ball's distance $x(t)$ increases indefinitely even after the ball hits the ground at $t_h$. Beyond missing friction from the ground, the ground doesn't even exist in the model at all. We can see this because the equation $y(t)$ will continue to decrease (at increasing rates) after $t_h$. However, because Alice only is modeling the ball's flight, the target of the model is simply the position of the ball when it is flying between time $t = 0$ and time $t = t_h$.

The target establishes a clear boundary for a model. Alice only cares about modeling the ball in flight, meaning she doesn't evaluate her model at times when that is not the case. After the impact at $t_h$, the above model stipulates the ball will continue to go into the ground at ever increasing speed. This is obviously not possible in the real world, but this doesn't threaten the model-target relationship because the target simply isn't the behavior of the ball after impact and the model performs well on the actual target of the ball's flight. Thus, targets can be very constrained, yet still have a justifiable model-target relationship.

If Alice instead wanted to consider a target world where the cannon was placed on the planet Mars, she'd have to change her model such that it represents this new target. Specifically, because we know that objects on the surface of Mars behave in much the same way to objects on Earth, she can use the same equation with $g$ changed to match the gravitational acceleration on the surface of Mars ($g_0 \approx 3.7m/s$). However, it is imperative that Alice know that the same kinematic equations apply on Mars, otherwise the model won't be a good one for the new target. While the change is small in this case, we could easily imagine a case where Martian objects didn't behave according to classical mechanics. For instance, consider a possible world where gravity on Mars varied significantly from one second to the next due to rapid movements in mass in the planet's core. Alice would need to account for this time dependence in her cannonball motion equations by adding terms to the model equations to ensure they actually replicate this different target.

Now consider what a simulation of this model might look like. In this case, such a simulation is relatively trivial as the equations are closed-form. Alice could simply convert the kinematic equations given above to a computer program, then run them at increasing time-steps $t = 0, 1, 2, \ldots n$. Using this simulation, she could reach a similar answer to the analytical solution reached above. When she notices that $y(t)$ dips below 0, she will know the modeled cannonball has hit the ground and can consult $x(t)$ to see how far the ball traveled. Notice that this answer will be approximate if $y(t)$ is zero at a fractional time, like 1.1s, which is an artifact of discretizing the continuous kinematic equations to discrete time values in the simulation.

## Purposes of Models

Different types of models can have a wide variety of purposes and a model that is a great model for one purpose might be a very poor model for another. For example, a globe is a physical model of the planet intended to help us learn about where things are on our planet, but a globe is a very bad model of earth's gravity. Thus, this model excels at being understandable and educational, but does not have much predictive power. On the other hand, equations from general relativity model time in relation to gravity well enough that engineers could predict precisely how much faster clocks on orbiting GPS satellites tick in orbit. Thus, relativity has great predictive power, but the equations themselves are certainly not as educational and immediately understandable to a layman as a globe is.

Given that

For the mathematical representational models described above, a model could be intended to improve understanding of

## Evaluating Models

For a model to faithfully model a target, that model must relate to the target by some relation $R$ which is defined over $\mathcal{M} \times \mathcal{W}$, where $\mathcal{M}$ is all possible models and $\mathcal{W}$ is all possible targets. Now, for some model $M$ and target world $W$, we find $R(W, M)$ if:

1. The modeler intends that each model part $m_i \in M$ is about a part of the target $w_i \in W$.
2. The math of $m_i$ must faithfully replicate certain properties of $w_i$ the modeler deems important.
3. Any claims the modeler makes about $W$ using $M$ must stem from properties that each $m_i$ replicates about each $w_i$.

These three conditions stem from the basic structure of an argument that uses a model, which is that of an argument by analogy. Essentially, it is on the modeler to show that the model is a reasonable analogue for the target and that the claims follow from the model or simulations run on the model. If the model and target have multiple parts, each model part must be shown to be a reasonable analogue for each target part.

This relationship between model $M$ to target $W$ need not be one-to-one. Consider if it were not a ball, but a Mars rock and if it were not thrown, but dislodged from a cliff by wind. We could still use the same model $M_b$ that we used to model $W_b$, but the target is now $W_r$, which is the world with the rock on Mars rather than the world with a ball on Earth. We know now that the same equations of mechanics work on both the Earth and on Mars with a modified acceleration due to gravity as we discussed earlier, but this isn't necessarily true. Physicists had to show that the model $M_b$, with all its parts, corresponds and faithfully replicated the behavior of $W_r$ even if $M_b$ is already know to faithfully replicated $W_b$. Otherwise, they couldn't simply use $M_b$ with modified parameters to make claims about $W_b$.

Say, for the sake of argument, that Mars has a very idiosyncratic core which shifts its mass around rapidly and continuously. All this movement would change the amount of matter below any object on a short timescale, meaning that the acceleration due to gravity on the surface would change as a function of time. In this case, $M_b$, while it worked great on Earth, would not work on Mars because acceleration due to gravity in the target $W_r$ is no longer constant, so $M_b$ would make bad predictions because it models gravity as a constant acceleration. Thus, $R(M_b, W_r)$ does not hold because a part of the model doesn't match with a part of the target. Given this mismatch, we can't trust that $M_b$ can be used to make claims about $W_r$.

In the case where a target is *hypothetical* the relationships between the model and the target might be trivial. In a video game, the mathematics in the model $M$ in some sense *define* the target game world $W$. The model $M$ will tell us quite a bit about $W$, but simply because the developer has defined $W$ as $M$.

Another example of these sorts of trivial relationships might be a modeler who is simply interested in the mathematics of $M$. In that case, the target is simply $M$ because the modeler is interested in the math itself, not the math as a method of standing in for some other target phenomena.

Before I move on to Zollman's models, I want to stress that this framework presented above is primarily to establish a clear and consistent way of naming and referring to models and the role they play in an argument. I don't posit that modelers themselves think like this when they are at work, simply that the above framework is a reasonable way to clearly get at some of the mechanisms at play when a modeler uses a model to make an argument about a target.

### Notation for Model Structure

The relationship between the target system and the machinery of the model that represents it can become a rich and complicated one, so it is worth developing a clear way of discussing this correspondence. These target-model relationships grow especially complicated in computational models which are often multifaceted and combine several distinct bits of model machinery.

To specify all this more precisely, let a model be specified as the set $M = \{ m_1, \ldots, m_n \}$, where an $m_i$ is a specific piece of model machinery. In a broad sense, these pieces of the model are separable portions of the model that can have their own target subsystems. This may mean an $m_i$ may be a variable, an equation, a piece of code or anything else that can be picked out and said to be standing in for some object in the target world. This division of the model can happen at varying levels of granularity. At a coarse level, the model $M_b$ of the ball might only include a single part $m_1$ which is the height of the ball over time $h(t)$. In this view, acceleration due to gravity, initial velocity and initial position are all implementation details and parameters for $m_1$. At a finer grain, the modeler might consider the position over time as model part $m_1$ that depends on separate model parts $m_2, m_3, m_4$ for gravity, initial velocity and initial position, respectively. These parts may interact with one another, as they do in the kinematic equation, because different parts of the target system often interact with each other. The point of these divisions is to allow a higher level of precision in specifying relations between the target system and the model, not necessarily to impose any constraints on how a model is internally divided.

There may be cases where a model does a great job at reflecting the behavior of some phenomena with extraordinary different internal structure than that of the target system. In such cases, a model may not need to be divided into parts. However, when modeling complex systems such a division can make it much easier to assess what the target system of a model is. For example, if the intended target system of a model is a complex biological system composed of many cell types, might be easier to see the correspondence between model and target if the model is divided into parts corresponding to different cell types. With such a division, if modeler can justify that each piece of the model represents its target cell type faithfully, they can have some faith the overall model will represent the complete target system. This method of justification of the relationship between model and target is especially intriguing when the emergent behavior of a target system is unknown, but the parts of that target are fairly well-understood.

Now, using similar notation as for the model, we can denote the target world as the set $W = \{w_1, \ldots, w_n\}$. Again, $W$ as a target system, may represent the actual world or a hypothetical world of the modeler's choosing. The modeler has significant latitude as to which target $W$ corresponds to her model $M$ because one mathematical model can be taken as having several different targets. Consider the thrown ball model $M_b$ we described above. We can say that its target $W_b$ is the ball and its behavior when thrown at some velocity and being effected by Earth's gravity. If a model is broken up into parts, each model part $m_i \in M$ will have some relation to some part of the target world $w_i \in M$. What exactly this relation is will depend on how the model is divided, the purpose of the model and even the overall strategy for evaluating the model.

Frigg, Roman, and Stephan Hartmann. "Models in Science." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2018. Metaphysics Research Lab, Stanford University, 2018.

Winsberg, Eric. "Computer Simulations in Science." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2019. Metaphysics Research Lab, Stanford University, 2019.