Homeostasis and Causal Control

Naftali Weinberger

May 12, 2025

Abstract

Homeostasis is a core biological concept concerning the way organisms maintain quantities within a given range. Mathematical models of homeostasis build on mechanisms for feedback control from engineering. The concept of control also figures centrally in causal analysis, where it is typically linked to interventions. This paper uses dynamic causal models (Iwasaki and Simon, 1994) to bridge these three domains, leading to the correction of a misunderstanding common in all of them. In each, control is regularly understood as aiming to maintain certain variables at constant values. Yet paradigm control feedback and homeostatic mechanisms do not maintain these variables constant, and intervening on them would destroy the mechanisms' functions. Instead of focusing how feedback loops allegedly maintain constancy, dynamic causal models reveal how control systems exploit such lower-scale feedback loops to produce higher-scale causal relationships. The discussion points toward a novel engineering-inspired approach to causal inference, and clarifies how biological and engineered systems combine existing mechanisms to perform increasingly complex functions.

1 Introduction

Homeostasis refers to the processes by which organisms maintain a quantity (such as body temperature) within a particular range. It is a core concept in biology, since it enables organisms to maintain stable internal conditions diverging from those of their external environments. While particular homeostatic mechanisms are complex and not always fully understood, *dynamic systems theory* provides insights into the conditions under which stability can be maintained. In general, a variable's stability requires a negative feedback loop by which it influences its own rates of change (given by time-derivatives). Within engineering, such stability-maintaining feedback loops are achieved through the design of *control feedback mechanisms* – a class of devices that maintain some quantity within a desired range. Paradigm examples are thermostats and cruise control mechanisms, which regulate room temperature and car speed, respectively. The modeling of control feedback is a common task in engineering, and provides a starting point for mathematically modeling homeostasis in biology.

A central motivation for studying biological mechanisms is to learn how to intervene to improve, protect, or restore their functions. This task falls squarely within the domain of causal analysis, since causal relationships provide bases not only for explanation and prediction, but also control. Unfortunately, despite great advances in causal modeling over the last 30 years, standard causal models cannot represent the feedback loops that maintain homeostasis. Such models typically do not allow for cycles, nor do they employ time-derivatives to capture a system's dynamics. For many decades there was only a trickle of work applying causal models to dynamical systems (Simon and Rescher, 1966; Iwasaki and Simon, 1994; Dash, 2003), though recently there has been an explosion of research causally modeling differential equations (e.g. Mooij et al., 2013; Rubenstein et al., 2016; Boeken and Mooij, 2024). In this paper I build on some of the older work, which has also been revived within the more recent literature (Blom et al., 2021; Blom and Mooij, 2023). In particular, I use the dynamic causal modeling framework of Iwasaki and Simon (1994) to model a paradigm control feedback system: the *Watt centrifugal governor*. The Watt governor was developed in the 19^{th} century to regulate the speed of steam engines. In addition to being a textbook example of a control system, the governor has also inspired a revolution in cognitive science due to van Gelder (1995), who used it to motivate an analysis of cognition not involving mental representations. Moreover, the mathematics for control systems such as the governor overlaps with that underlying reinforcement learning methods for unsupervised machine learning, which are now prominent within artificial intelligence.

The process of causally modeling the Watt governor reveals that prior discussions of its differential equations all downplay key features that are essential to its function. These discussions ubiquitously emphasize its use of negative feedback to maintain the speed of an engine at (or at least near) some desired value. But the governor does not, in fact, maintain engine speed constant, and its use of negative feedback does not differentiate it from simpler stabilizing processes. The governor does facilitate engine stability, though here I will emphasize its further function of adjusting steam supply to meet the engine's varying energy demands. This feature of the governor's behavior is obscured within standard discussions of its dynamics, but emerges clearly in the causal model for the governor at equilibrium. It is essential for linking the governor's higher- and lower-scale dynamics, and thus for more generally understanding the role of controllers as components in more complex systems.

The subtle shift from thinking of the governor as a device for maintaining a quantity fixed to instead seeing it as matching energy supply to meet varying demand has massive implications for homeostasis. Like engine speed, quantities such as body temperature and blood pressure that organisms allegedly maintain constant in fact fluctuate to meet varying energy needs. To highlight this, Sterling (2004) proposes supplementing the concept of homeostasis with that of *allo*stasis, which emphasizes the longer-term mechanisms by which organisms anticipate and coordinate the varying energy demands of their subsystems. Whereas discussions of homeostasis and self-regulating feedback loops often fetishize the emergence of order in the absence of centralized control, allostatic processes coordinate competing energy demands via the central nervous system. While the governor is simple in comparison to biological regulatory mechanisms, even this simple device is better understood as playing the energy-matching role of allostasis than the value-fixing role often attributed to homeostasis.

The discussion undermines the common identification of causal control with interventions on variables. While standard "arrow-breaking" interventions fix a variable's value independently of its prior causes, the governor's dynamic causal model reveals that control systems require one *not* to intervene on variables such as engine speed. Accordingly, discussions of the governor emphasizing fixing over matching miss that it could not play its matching role were engine speed to be fixed via intervention.

The modeling approach adopted here points towards a novel form of causal inference that is inspired more by engineering than computer science. Despite the cliche that causation is about control, existing causal models are passive: they are used to represent *found* systems, rather than to *construct* systems supporting desired causal relations. Interventions uncover the properties of existing systems rather than being used to create novel ones. In contrast, dynamic causal models uncover how engineered and biological systems exploit lower-scale feedback loops to bring about higher-scale control, thus providing insights into how one might build mechanisms enabling causal control.

2 The Watt Governor as a Model for Homeostasis

Steam engines were developed in the 18^{th} century, and were a crucial component of industrialization. Yet their widespread adoption required first resolving an engineering difficulty. For many tasks, the engines needed to rotate at a constant rate, but engine speed depends on the steam fed into the engine and the workload, and both of these inputs are variable. For instance, an engine might be used to power multiple looms, and attaching an additional loom would increase the workload. But one would not want this addition to slow down the engine and disrupt the activity of the looms already running. What is therefore required is a way to adjust the steam being fed into the engine such that the steam flow is increases when the engine slows down and decreases when it speeds up.

One might imagine this problem calls for a worker whose job it would be to let in more steam as the engine slows down and less when it speeds up. But such a worker might not react quickly enough to make the necessary adjustments. The Watt governor is an elegant device that makes the adjustment automatically.[1] It employs a vertical spindle that rotates at a rate proportional to the rotational speed of the engine, ω. The spindle is attached to two flyballs on two arms at angle θ (fig. 1(a)), and the arms are hooked up to a sleeve that determines how much steam is fed into the engine via a throttle valve (fig. 1(b)). To see how this works, suppose that the engine accelerates above its desired speed. The increase in speed will lead the arms to go outward due to centrifugal force, and as the arms go outward the throttle valve will adjust to allow in less steam. In short, increases above the desired

[1] It perhaps bears mention that dissatisfaction among textile workers with the results of automatization is what led the Luddites to break the new machines in protest.

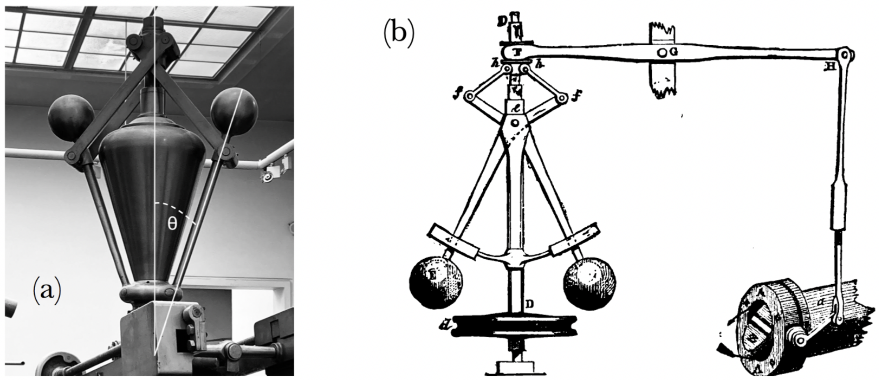


Figure 1: (a) Photo of a Watt Governor indicating angle θ; (taken at the Deutsche Museum) (b) Governor and Throttle Valve

speed lead to less steam and decreases lead to more, thus keeping the engine's speed constant despite changes in workload.

So far I've presented the standard account of the governor's function. Beltrami (1987), for instance, presents the governor as keeping the "engine running at a constant angular speed in the face of varying loads" (p. 162, cf.: van Gelder (1995, p. 347), Munis (2012, p. 142), Billman (2020, p. 5), Sterling (2020, p. 39)). This, however, turns out not to be accurate, or even consistent with Beltrami's own derivations, which reveal that engine speed is not constant, but depends on workload. The dependence of engine speed on workload is in fact relatively weak, so more careful claims that the governor maintains speed within a limited range (e.g. Pontryagin, 1962, p. 217) are not incorrect per se. But even these focus on constancy in a way that I'll argue makes it harder to understand its function.

My primary motivation for discussing the Watt governor is as a basis for modeling homeostasis. Cannon (1929) coined the term "homeostasis" as a way of capturing Bernard's (1865) insights about organisms' need to maintain their internal environments within a stable equilibrium state that remains nearly constant despite external changes. Common examples of quantities regulated by homeostasis are body temperature, arterial blood pressure, blood sugar, and extracellular pH. The link between the governor and homeostasis is not new to this paper. The governor is sometimes used to introduce homeostasis (e.g. Billman, 2020; Munis, 2012), and Sterling (2004, p. 26, fn. 3) even speculates that Bernard himself was inspired by the development of steam engine governors.

The governor is a paradigm example of a control feedback mechanism. The mathematical treatment of such mechanisms is commonplace within engineering (Franklin et al., 2002). So any generalizations of the governor's model to more complicated systems is likely to fall within well-explored mathematical territory. As an added bonus, the discussion here is likely to have implications for applications related to control theory, including reinforcement learning and Kalman filters (Farsinezhad, 2015).

Norbert Wiener's *cybernetics* project uses control feedback as a model for cognition, and, unsurprisingly, he discusses the governor (Wiener, 2019, p. 97). Within more recent cognitive science, the governor became well known through van Gelder's (1995): "What might cognition be, if not computation?". There he employs dynamical systems theory to offer an alternative to computational analyses of cognition, on which problems are solved by dividing a task into sub-tasks, which are then performed by specialized modules. Crucially, while the computational approach would suggest designing a specialized module measuring the speed of the engine and a separate module adjusting the steam flow, the Watt governor does not do this, and, per van Gelder, should not be understood as containing a representation of engine speed. This serves as an analogy for how cognitive tasks may be performed without mental representations. The governor is thus central to ongoing debates in cognitive science.

Within philosophy of science, it is almost a cliché to say that causation matters for "explanation, prediction, and *control*". So one might have naively assumed that applying causal methods to control feedback systems would be straightforward. Yet standard causal models apply to static systems without feedback, so generalizing them to devices such as the governor is non-trivial, and will take up a substantial proportion of what follows. Fortunately, this work will have significant payoffs. The process of causally modeling the governor will reveal insights that have been neglected in the 150 years of its mathematical analysis, and these insights matter when using it as a model for homeostasis.

3 Background: Graphical Causal Models

Causal models typically employ *directed acyclic graphs* (DAGs) in which the nodes are random variables and the arrows denote direct causal relationships. "Acyclic" indicates the absence of causal loops – i.e. one cannot get from a variable back to itself via a set of connected arrows all going in the same direction. *Structural equations* present the functional relationship between each variable and its direct causes, indicating how the effect variable would change given interventions on its causes. In cases in which one can adjust for confounding, it is possible to learn these structural equations from the DAG and the joint probability distribution over its variables.

Causal DAGs are employed for a range of distinct projects. *Causal discovery* refers to the use of algorithms to choose among candidate causal hypotheses given a joint probability distribution (Spirtes et al., 2001). *Causal identification*, in contrast, assumes a given causal model and provides criteria for when a particular causal quantity (e.g. an average effect) can be derived via formulas adjusting for the influences of confounding factors (Pearl, 2009). Causal identification is closely connected to the notion of an intervention, since in cases where one can identify a structural equation from the distribution, one learns how the effect would respond to hypothetical interventions on its causes, even without actually intervening. Yet experimental interventions enable one to identify the effect even when it is not identifiable from observational data given one's assumptions. They do so by changing the causal structure such that the intervened-upon variables are no longer influenced by confounders.

The most prominent philosophical application of graphical causal models has been in developing accounts of causal explanation. Notably, Woodward's (2003) interventionist account of explanation differentiates causes of effects from mere correlates based on the fact that interventions on causes influence their effects. Here the notion of an intervention can be formally defined within a causal model, and is sometimes represented as breaking the arrows going into the intervened upon variable. Unlike philosophical accounts that seek to reduce causal claims to claims about some allegedly more basic ontological category (e.g. probabilities, counterfactuals, or processes), Woodward's account is *non-reductive*. This means that it systematically relates causal claims to claims about probabilities and counterfactuals, but always relies on causal assumptions when drawing causal conclusions. For instance, one cannot know whether an action counts as an intervention without making causal assumptions. Note, however, that such assumptions do not require that one actually intervene on a system in order for the notion of an intervention to be well defined (Zhang, 2022). Although (and perhaps because) Woodward does not pursue the traditional reductive project, interventionism has become the dominant account for understanding causal explanations across the social and life sciences.

The focus of this paper is not on any particular application of causal models, but rather a generalization of the graphical framework that underlies all of them. Graphical causal models have traditionally been applied to static systems (Weinberger, 2019), and cannot represent the differential equations used in control systems theory. There has recently been a tremendous amount of recent work bridging the gap between differential equations and causal models (e.g. Mooij et al., 2013; Rubenstein et al., 2016; Boeken and Mooij, 2024). The older Iwasaki and Simon (1994) framework I adopt here is specifically suited for modeling the dynamics of systems that are perturbed away from equilibrium (See Blom et al., 2021; Blom and Mooij, 2023, for recent refinements). Because control feedback and homeostatic systems maintain stability in light of such perturbations, this framework is a natural starting point for modeling them. Additionally, it is philosophically advantageous that Iwasaki and Simon's framework is a generalization of Simon's earlier causal ordering method (Simon, 1953), which provided conceptual foundations for the DAGs and structural equations mentioned above.

4 Dynamic Causal Models

This section introduces dynamic causal models using an example. Those interested in further technical details may consult the cited literature, including: Simon and Rescher (1966); Iwasaki and Simon (1994); Dash (2003); Weinberger (2020, 2021); Blom et al. (2021) and Blom and Mooij (2023).

Simon's (1953) causal ordering method takes a set of standard equations, and derives a set of structural equations in which each variable is given as a function of its causes. To illustrate, the ideal gas law relates a gas' pressure (P), temperature (T), and volume (V) at equilibrium:

$$PV = kT \tag{1}$$

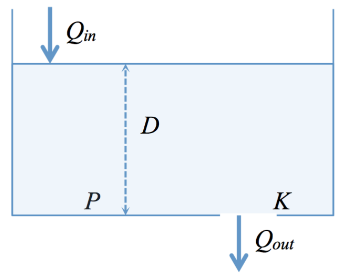


Figure 2: Bathtub with water flowing in at rate Q_{in}, out at rate Q_{out}, with depth D, pressure P and drain size K

In equation (1), which variables are placed on which side of the equals sign is just a convention – algebraically shifting variables to either side does not change its content. Consequently, the equation provides no information about whether, for example, T causes V, V causes T, or neither. Yet, in combination with further information about the system in which the variables are instantiated, we can use the gas law to derive causal directionality. If the gas is in a fixed-volume container in a heat bath, we can add equations indicating that the values of temperature and volume are exogenously fixed:

$$T = t \tag{2}$$

$$V = v \tag{3}$$

Simon's causal ordering method implies that because the values of T and V need to be solved in order to solve for the value of P in (1), T and V cause P (graphically: $T \rightarrow P \leftarrow V$). Alternately, were the gas in a movable piston instead of a sealed container, we would replace (3) with:

$$P = p \tag{4}$$

Since one needs to solve for T and P in order to solve for V, temperature and pressure cause volume. More generally, the causal ordering of a set of variables (i.e. what causes what) corresponds to the order in which one solves for the variables in a set of equations.

When it is possible to solve for the causal ordering, one can rewrite the original equations as a set of structural equations. For example, given (1), (2), and (3), we can rewrite (1) as:

$$P = \frac{kT}{V} \tag{5}$$

By convention, the variable on the left-hand side is an effect of those on the right (despite the misleading "="). Nowadays, it is standard to introduce equations with the causal directionally already specified, as in (5). Simon's method shows how to derive such asymmetric ("structural") equations from a set of symmetric ones. To be clear, "derive" here does not mean "get causal knowledge without causal assumptions". It should already be clear that the causal equations one ends up with depend on the equations one begins with, and those build in causal assumptions about the system – most saliently which variables *don't* depend on the others, and are thus *exogenous*. Nevertheless, the causal ordering method helps clarify that the core assumptions of causal modeling are independence assumptions (Hausman, 1998), since the possibility of a unique ordering depends on not every variable appearing in every equation. Additionally, note that causal ordering does not depend on the time-ordering – the procedure did not rely upon the timing of the variables, which in the example are all at equilibrium.

Iwasaki and Simon (1994, p. 145) note that Simon's method was originally developed to model static systems in which the variables are measured at equilibrium or steady state. To illustrate their generalization of the method to systems that away from equilibrium, I'll walk us through their example involving a bathtub into which water flows in and out. Given the complexity of the systems covered later in the paper, a humble bathtub might not seem like a promising starting place. Yet the bathtub exhibits a form of long-term feedback in that it has an equilibrium point at which its rate of flow out

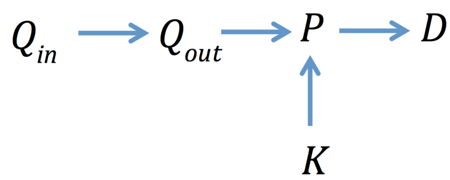


Figure 3: Causal Ordering for the Bathtub at Equilibrium

equals the rate of flow in, and the system resists changes away from this equilibrium.[2] If one were to change the rate of inflow or outflow by turning a faucet or widening the drain, the depth of the water will change until the outflow once again equals the inflow (assuming the tub does not overflow or fully drain). Despite the bathtub's familiarity, the process of modeling it requires subtle attention to its longer- and shorter-term dynamics, as well as the linking of variables representing stocks and flows.

The bathtub's variables are as follows (fig. 2). Water flows in at a rate of Q_{in} and out at rate Q_{out}. The rate at which the water flows out depends on its depth D, the resulting pressure P, and the size of the drain, K. Intuitively, the difference between Q_{in} and Q_{out} determines the rate of change of depth, which determines future D and P, which, along with K determines Q_{out}. Yet the system also has longer-term term dynamics, since if it neither fully drains nor overflows, there will be an equilibrium point at which $Q_{in} = Q_{out}$.

We can write the equations for the bathtub at equilibrium as follows:

$$0 = f_6(Q_{in}) \tag{6}$$

$$0 = f_7(K) \tag{7}$$

$$0 = f_8(P, D) \tag{8}$$

$$0 = f_9(K, P, Q_{out}) \tag{9}$$

$$0 = f_{10}(Q_{in}, Q_{out}) \tag{10}$$

The equations are written without the functional form specified and with all of the variables on the right-hand side in order to emphasize that we have not yet solved for the causal ordering. Equations (6) and (7) indicate that the rate of flow in and the size of the drain are exogenous. (8) indicates that pressure and depth are proportional, and (9) indicates a relationship between pressure, the size of the drain, and Q_{out}. Finally, equation (10) indicates a relationship between Q_{in} and Q_{out} at equilibrium, since at equilibrium these two rates need to be equal.

Using the causal ordering method, one derives the graph in figure 3, which should seem a bit counterintuitive. It might seem obvious, for example, that Q_{in} causes Q_{out} *via* P and D. Here it crucial to bear in mind that we are considering the system only at equilibrium. In the short run, changes to the rate of Q_{in} will lead to *transient* changes in depth and pressure as the system reaches equilibrium (which, if the change to Q_{in} is not itself transient, will be a *new* equilibrium). But the only thing that the equilibrium value of Q_{out} depends on is Q_{in}. If, for example, one tried to change Q_{out} by changing K, this would influence equilibrium P and D, but would have no long-term effect on Q_{out}. Although further opening the drain would initially lead to an increased flow, as the water drains the depth and pressure will go down until Q_{out} once again equals Q_{in}.

Given the timescale at which we usually observe bathtubs, we get a more intuitive causal ordering when considering the system away from equilibrium. The key step is replacing equation (10) indicating that $Q_{in} = Q_{out}$ with:

[2] Historically, systems employing self-regulating tubs, such as Ctesibius' water clock in 4th century BC Greece, can be seen as the origin of control feedback engineering (Kang, 2016).

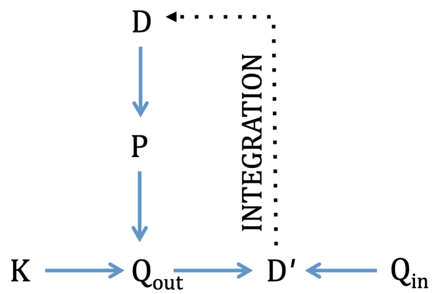


Figure 4: Dynamic Causal Ordering for the Bathtub

$$D' = f_{11}(Q_{in}, Q_{out}) \tag{11}$$

Equation (11) indicates that when the system is away from equilibrium, the rate of change of depth (given by its time-derivative D') depends on Q_{in} and Q_{out} (in fact, it equals $Q_{in} - Q_{out}$). In principle, one could also include time-derivatives for every variable in the model, so the model we will consider is in fact a "mixed" rather than a purely dynamical causal model. This amounts to considering the system at a timescale at which D has not yet reached equilibrium, but the other variables have had time to adjust to any perturbations of their causes, thereby assuming that not every variable equilibrates at the same rate.

Including a time-derivative for D in the model requires adding another equation:

$$D_{t+1} = D_t + D'\Delta t \tag{12}$$

Equation (12) involves a discrete version of the mathematical operation of integration, which takes D' and D at time t and yields D's value at a subsequent time-step $t + 1$. A textbook example of this would be using an object's velocity and position to determine its position at a later time. The fact that one can only predict D's future values given its current value is significant. It means that in dynamic causal models, unlike standard ones, one needs to specify initial conditions. This matters for the causal ordering, since we treat D as exogenous.

After replacing equation (10) with equations (11) and (12), we can derive the *dynamic causal model* (DCM) in figure 4. (11) indicates that D' depends on Q_{in} and Q_{out}, and the dashed arrow corresponds to the derivation of D from D' using integration (i.e. 12). Since D is now treated as exogenous, the rest of the causal ordering can be straightforwardly solved such that D causes P, which, along with K, determines Q_{out}. The DCM thus captures a much more intuitive ordering of the system.

Within dynamical systems theory (Strogatz, 2024), it is well-known that negative feedback loops are crucial for producing stability. Dash (2003) presents a *structural stability criterion* proving that for DCMs to reach a stable equilibrium it is necessary that every variable whose derivative is included in the model has a causal (possibly indirect) influence on its higher-order derivatives. This coheres with a wide range of stability conditions in the mathematical literature, and means that DCMs provide a basis for quickly determining whether a necessary condition for causal stability is satisfied.

Iwasaki and Simon (1994) provide an operation called *equilibration*, which allows one to derive an equilibrium model from a dynamic one. Dash (2003) provides a schema for equilibrating X:

1. Set all derivatives of X in the model to 0 and remove them from the model
2. Delete all equations going into X or its derivatives
3. Remap to get the new causal ordering

In our example, Step 1 replaces D' in equation (11) with 0, transforming (11) into (10). Step 2 deletes (12). Solving for the causal ordering in Step 3 yields the equilibrium graph.

Equilibration can be thought of as "zooming out", in that one starts at a timescale at which a variable has not had time to reach equilibrium, and shifts to a timescale at which the equilibration time is negligible. Iwasaki and Simon also present an *exogenization* operator, in which one zooms *in* to a time scale at which a slowly changing variable changes so slowly that one can treat it as fixed. If we were living in a slow-changing "bathtub world" in which we never reached equilibrium, and there was a permanent gradient between Q_{in} and Q_{out}, we might take it as a basic fact about our universe that

D causes P, without being aware of the longer-term feedback loop. This might seem strange when considering the bathtub, but perhaps seems less strange when one considers accounts of the direction of time suggesting that it depends on a global entropy gradient going back to the beginning of the universe (Albert, 2003; Ismael, 2023). When all of our observations of a system are at a timescale at which it is in the process of equilibrating, we don't think of equilibrium as a temporary perturbation, but might mistake the properties of the equilibrating process for laws of nature.

Causal graphs derived via equilibration have a worrisome property that has hindered their more widespread adoption. Namely, they break down under interventions on certain variables. This occurs when such interventions destroy feedback loops necessary for the system to reach equilibrium. In the bathtub, for instance, if one were to intervene to hold D fixed to a particular value – say, by introducing a different water source that compensates for any change in depth – then Q_{out} would no longer match Q_{in} at equilibrium. This seems to contradict the equilibrium model, in which Q_{out} is causally before D, and thus presumably should not be influenced by interventions on it. Blom et al. (2021) use the bathtub to systematically explore the way that equilibrated models succeed and fail at predicting different types of interventions, concluding that models such as the equilibrium bathtub model "lack any obvious causal interpretation" (39). This resembles an earlier conclusion by Dash (2003), who provides the conditions under which equilibration will produce such problematic models, and argues that equilibrium models will not generally be reliable.

Weinberger (2021) responding to Dash (2003) and Dash and Druzdzel (2001), defends the causal interpretability of equilibrium models. He does so by rejecting a condition, which Dash and Druzdzel (2001) label the *manipulation postulate*, requiring that adequate causal models be valid under interventions on *all* variables in the model. Weinberger's rejection of the postulate reflects that equilibrium causal models apply only under conditions in which the systems is presumed to reach a certain equilibrium state, and interventions preventing the system from reaching that state transform it to a different one to which the model no longer applies. Given a substantive understanding of how the bathtub functions, it is unsurprising that Q_{out} will not equilibrate to match Q_{in} if one holds D fixed, since holding D fixed requires that Q_{in} does not, in fact, include all flows of water into the tub.

Putting aside abstract issues of model adequacy, it *is* worrisome that certain equilibrium models fail to predict certain interventions, without indicating this limitation. One of Dash's concerns is that if one uses a constraint-based causal discovery algorithm (e.g. PC) to infer the causal model from equilibrium data, one will only have access to the equilibrium model, and thus have no way of knowing the conditions under which the equilibrium model breaks down. The bathtub equilibrium model alone provides no indication that an intervention fixing D or P's value would in fact change Q_{out}. Fortunately, this concern will not affect what follows. Here the question is not when one can use the equilibrium model in the *absence* of the corresponding dynamic model. Rather, I'll be starting with the dynamic model, and be concerned with whether the model derived from equilibration reveals important information that would be missed from looking at the dynamical model alone. This helps, because one can still use the dynamic model to specify the variables in the equilibrium model that one cannot intervene upon without destabilizing the system. Subject to this limitation, I will argue that the equilibrium model is still indispensable for predicting the effects of certain longer-term interventions.

In the dynamic model for the bathtub, D is the only variable that regulates itself (that is, causes its own derivative). Fixing the value of any variable on the self-regulating loop – in this case, D, P, or Q_{out} – will destroy the ability of the system to equilibrate, and the equilibrium model will not remain valid under interventions of any of those variables. But, subject to this limitation, the model is still informative. It tells us that if we want to change equilibrium Q_{out}, we can do so via changing Q_{in}, but not via changing K. This is not what one would expect from looking at the dynamic causal model, in which K directly causes Q_{out}. Additionally, the model tells us that one can change equilibrium depth and pressure either by changing Q_{in} or K.

To briefly illustrate how the bathtub example generalizes, let's shift our perspective away from what happens inside the bathtub, towards the function it might play as part of a broader system. Imagine a series of bathtubs in which the outflow of each is the inflow of the next. An intermediate bathtub might act as a reservoir so that water could be quickly released into the downstream tub when necessary. The equilibrium model would then tell you how to adjust the size of the drain in order to have the desired amount of stored water. In such a system, the function of a bathtub would be similar to that of a capacitor in an electric circuit, where one is primarily interested in equilibrium scenarios in which a current is flowing, and where we observe the system at a timescale at which it is natural to

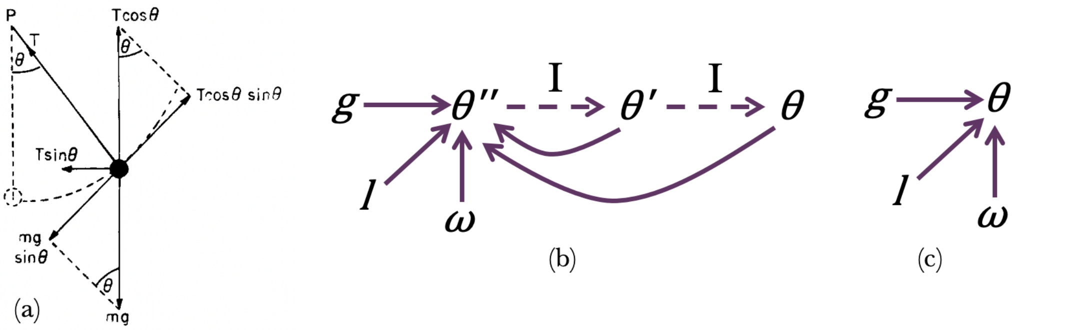


Figure 5: (a) Direction of forces for a rotating pendulum (b) dynamic model (c) equilibrium model from equilibrating θ

treat the capacitor as a black box. The bathtub thus serves as a model for other systems maintaining constant flow, where the equilibrium model abstracts away from the flow-maintaining mechanism.

5 Causally Modeling Feedback Control

The mathematical analysis of the governor is over 150 years old. In the anglophone world it is common to attribute it to Maxwell (1868), though Pontryagin (1962) credits Vyshnegradsky (1876) with providing a more rigorous treatment. Here I follow the textbook presentation of Beltrami (1987), which is very similar Pontryagin's. Given this long history, one would expect that there is little to be discovered about the governor's mathematics. Yet in what follows, I'll argue that the dynamic causal model for the governor illuminates essential features of its function that are obscured within standard mathematical treatments.

Weinberger and Allen (2022) informally present a DCM of the governor, though here I improve upon it and provide it with a more rigorous basis. Building a DCM for the governor is not especially tricky, as most of its content derives from the governor's two differential equations: one for the dependence of the flywheel arms' angle, θ, on engine speed ω, and one for the way that ω depends on the steam fed into the engine and the workload. It is also crucial that the amount of steam fed into the engine depends on the angle of the arms. But this is not given by a *differential* equation, as we'll see.

The equation for the dependence of θ on ω is the same as that for a rotating pendulum hanging from a hinge (Beltrami, 1987, pp. 152-5 example 6.5), so we begin by modeling that system:

$$\frac{d^2\theta}{dt^2} = \underbrace{(n\omega)^2 cos(\theta) sin(\theta)}_{\text{(i)}} - \underbrace{\frac{g}{l} sin(\theta)}_{\text{(ii)}} - \underbrace{r\frac{d\theta}{dt}}_{\text{(iii)}} \tag{13}$$

All forces acting on the pendulum's bob are given by vectors running tangent to the curve traveled by the pendulum of length l (figure 5(a)). Term (i) in the equation indicates that the horizontal centrifugal force due to the rotation of the pendulum ω is directed outward along the tangent (we'll ignore the constant n). Term (ii) indicates that the downward force due to gravity (g) is directed inward along the tangent. Term (iii) represents the resistance needed to ensure that the pendulum comes to rest at equilibrium, but will not be otherwise important here.

The DCM for this subsystem is given in figure 5(b). In addition to equation (13), one needs to further specify that (A) the variables g, l, and ω are independently given (i.e. are *exogenous*), and (B) when one derives a variable from it's higher-order derivative via integration, this requires specifying that variable's value at the prior time-step. Concretely, this corresponds to the following equations:

$$G = g \tag{14}$$

$$L = l \tag{15}$$

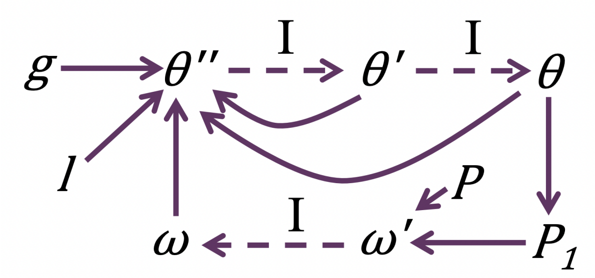
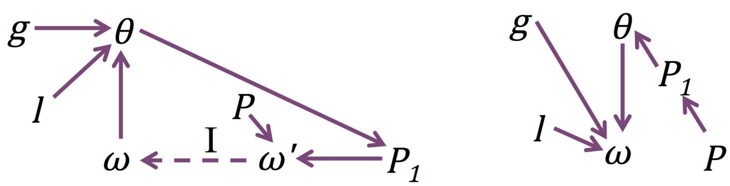


Figure 6: Dynamic Causal Model of the Watt Governor



(a)

(b)

Figure 7: The models resulting from equilibrating (a) θ and then (b) ω in the dynamic causal model.

$$\Omega = \omega \tag{16}$$

$$\theta_t = \theta_{t-1} + \theta' \Delta t \tag{17}$$

$$\theta'_t = \theta'_{t-1} + \theta'' \Delta t \tag{18}$$

Equilibrating θ (twice) yields the model in figure 5(c), which simply collapses the dynamical feedback loop without substantially changing the model.

In the governor system, ω is not constant, but rather depends on both the steam let into the engine and its workload. So we need to introduce an equation for ω', but even once we do so, we will not need to alter equation (13) to reflect that ω changes. Recall that when one derives ω via ω' using integration, one also needs to specify an initial value for ω, which will still count as exogenous within the dynamic causal model. The rotational formulation of Newton's second law gives ω' as a function of its component forces, namely: (1) the torque due to steam P_1, and (2) the torque in the opposite direction due to workload (P). More precisely:

$$I\omega' = P_1 - P \tag{19}$$

I refers to the rotational inertia of the flyballs. The value of P is exogenously given and P_1 depends on θ, since the throttle valve depends on the angle of the arms. These details are sufficient to derive the dynamic causal model given above in figure 6, which illustrates how ω regulates ω' via a feedback loop involving the governor.[3]

One needs to equilibrate θ prior to ω in order for equilibration of ω to be well defined. If you tried to equilibrate ω first, the system would be under-constrained, since after equilibration ω is not determined by an exogenous initial condition, and θ'' can only be calculated given ω. Equilibrating θ and then ω yields the equilibrium model in figure 7(b). This graph indicates that P determines P_1 at equilibrium (since according to (19) these must be equal when $\omega' = 0$), which in turn determines equilibrium θ and ω. From the dynamic graph, we know that only P, g, or l are suitable intervention targets – intervening on other variables will break the feedback loop necessary for stability.

[3]More rigorously, the required additional equations are: (22) $P = p$, (23) $0 = f_{23}(P_1, \theta)$, and (24) $\omega_t = \omega_{t-1} + \omega' \Delta t$. See §6 for the details of the θ-P_1 relationship in (23).

The equilibrium causal graph reveals that ω depends on the workload. Given that the governor is ubiquitously described as maintaining a constant speed despite changes in workload, this might appear to be a mistake. Yet the equilibrium solutions to the differential equations reveal that equilibrium ω *does* inversely depend on workload P (Beltrami (1987, p. 187), Pontryagin (1962, p. 219)). This dependence is no mere mathematical curiosity, since it took decades to develop control systems that could in fact maintain a quantity constant at a desired value. Whereas the Watt governor is an instance of *proportional control*, which cannot do so, *integrated controllers* can. It is the latter that are required for cruise control to maintain a car at constant speed, which was achieved in the 1920s (Ball, 2003).

The dependence of ω on P is not very strong, since ω inversely depends on the *square root* of P. So discussions that treat the governor as maintaining constant speed are not wide off the mark. But for the purposes of understanding the governor's function, the emphasis on constancy is misleading. For one thing, if the governor *did* maintain constant speed, it could not function: since θ both serves as a sensor measuring engine speed and as an activator adjusting the throttle valve, the device could not let in different amounts of steam without differences in speed. Additionally, now that we know that the governor does not maintain constant speed, we need to get clearer on what it in fact *does*. Part of the story is clearly that it uses negative feedback to keep speed *stable*, with different stable values for different workloads (note that stable $\neq$ constant). But even the simple pendulum model characterizing the flyballs involves a stability-maintaining negative feedback loop. So the appeal to negative feedback is less illuminating than one might suppose.[4]

The intuitively elegant feature of the governor is the way that it adjusts steam supply to match the system's varying energy demands – different workloads require different amounts of steam. This role is not obvious from the differential equations, but is clearly represented in the equilibrium causal model by the fact that P causes P_1. This is a longer-term relationship, akin to the influence of Q_{in} on Q_{out} in the bathtub system. In the same way that the bathtub system exploits the shorter-term feedback loop to achieve a long-term stable relationship between Q_{in} and Q_{out}, the dynamic causal model for the governor reveals how the system exploits the governor's feedback loop in order to produce a long-term stable relationship between P and P_1.

With this newfound appreciation of the $P \to P_1$ relationship, let's revisit the governor's DCM. To understand the governor, the dynamics are relevant not solely because of the feedback loop by which ω regulates itself. Equally important is how the governor uses this loop as a means by which P can be used to control P_1. This control relies on the facts that (1) P is exogenous, and (2) because P and P_1 cause ω', they must be equal when ω equilibrates. The equilibration condition that $\omega' = 0$ only ensures that ω has a constant equilibrium value *given* a specific workload, not that ω's equilibrium values should change little given different workloads. How much ω changes depends on how long it takes for it to equilibrate. Since steam flow equilibrates relatively quickly in response to changes in workload at the timescale of interest, ω does in fact vary little in response to changes in P. But from the perspective of the $P \to P_1$ relationship, this is incidental.

Contrary to the dominant analysis, the governor does not keep engine speed constant given changes in workload. What *is* being held constant is the higher-order relationship between workload and steam flow at longer timescales. Put differently, the governor should be understood not as *fixing* a quantity, but as creating a *matching* relationship. The difference is represented schematically in figure 8. Whereas fixing emphasizes the use of a feedback loop to keep a quantity fixed, matching refers to the use of that feedback loop to create a stable higher-scale relationship.

Given how sensitive the causal ordering method is to the specification of exogenous variables, one might wonder what this depends on. In some cases, exogeneity appears to be truly user- or modeler-dependent (Janzing and Mejia, 2024). For instance, although it is natural to say that a cyclist exogenously influences the bicycle's pedals, which cause the wheels to rotate, one could turn the bike over and rotate the wheels to influence the pedals (Druzdzel and Van Leijen, 2001). In other cases, exogeneity reflects features that are fixed at the spatiotemporal scale at which one observes a system, such as climate being treated as exogenous for agriculture over shorter scales (Weinberger, 2020). In the bathtub example, the direction of gravity is clearly relevant, and this direction would presumably remain a constraint as one incorporates the tub into a broader system. What matters here is that lower-scale equilibration processes give rise to higher-scale stable dependency relationships,

[4] This is not to deny that designing stable governors is much more difficult than designing stable pendulums, nor that the mathematical theory of stability spearheaded by the governor was a massive achievement. Nevertheless, the justified enthusiasm for theorizing about stability led to a neglect of other features of the governor's function.

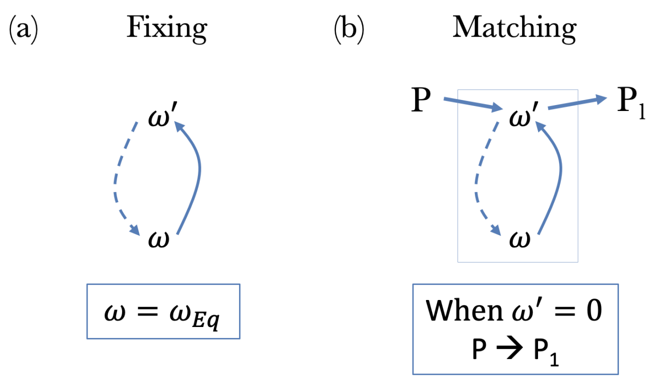


Figure 8: A schematic illustration of fixing vs. matching

which may then be exploited by engineers or evolution. The conditions that determine how such symmetric relationships can be exploited in a particular causal direction call for future research.

Here I've sought to de-emphasize the way that the governor keeps engine speed within a limited range. My aim in doing so is not to deny that this is a desirable feature, but rather to make clear why it is not the whole story. Both the governor's higher-scale steam-regulation function and lower-scale speed regulation function are relevant to understanding its operation. More generally, any complex system, whether engineered or biological, needs to satisfy constraints over multiple scales in order to succeed. Yet the higher-scale matching function of the governor has been severely neglected, which is why I've here been at pains to show how it is made salient within the governor's causal models.

6 What Gets Lost in the Mathematics

The idea that the governor adjusts steam supply to meet the engine's energy needs is not itself new. It is noted in many presentations of the governor, and I suspect that this elegant adjustment is what initially impresses people when they learn about the device. What has been lacking in prior discussions is not so much an understanding of what the governor does, but rather a proper link between our intuitive understanding of the governor and its mathematical analysis. This matters especially when using the governor as a model for thinking about other systems.

When engaging in a project involving translation from mathematical to less technical contexts, one opens oneself up to a criticism along the following lines: "Even if *many* misunderstand X, *true experts* already knew what you're saying." Here "experts" invariably means: the two people I trust on this topic, who have been studying it since 1983. The version of this response for the governor points out that it is clear from the differential equations that the governor does not ensure the *constancy* of ω, but rather its *stability* (i.e. that $\omega' = 0$ at equilibrium). As long as one is careful that the governor ensures stability rather than constancy, one avoids misunderstandings. Naturally, informal language will fail to capture the distinction between constancy and stability, but that's why we need mathematics. So what's the big deal?

To preempt this response, it is worth spending a bit more time to illustrate how far out of its way the standard mathematical treatment goes in order *not* to discuss the relationship between P and P_1. Readers not interested in the details are free to skip to the next section, but I'd suggest that the details are illuminating, and the clarification I provide is both novel and necessary.

My discussion here continues to follow the treatment of Beltrami (1987). Confusingly, Beltrami describes P as the "torque due to the variable load on the flywheel" (Beltrami, 1987, p.163). Yet P is static, and variation in P is reflected only in the influence of this variation on other variables in the system. More precisely, he gives the following equation:

$$P_1 = \bar{P}_1 + \alpha(cos\theta - cos\theta_0) \tag{20}$$

$\bar{P}_1$ corresponds to the torque due to steam at the "desired rotational speed ω_0" (p. 163) and θ_0 is angle of the arms at that speed. $cos\theta$ corresponds to the height of the sleeve corresponding to arm angle θ. So equation (20) reflects that the deviation of the amount of steam from the desired amount $(P_1 - \bar{P_1})$ is proportional (by constant α) to the deviation θ from its desired value. I'll soon say more about how to interpret $\bar{P_1}$ and θ_0, but for the moment the key points are: 1) at equilibrium ($\omega' = 0$) there is no deviation of P_1 from $\bar{P_1}$, or of θ from θ_0, and 2) deviations away from equilibrium correspond to a mismatch between P and P_1. So if P starts at a value at which the system is at equilibrium, and shifts to a value to which the system has not yet adjusted, this will not be reflected though any notation related to P, but rather through P_1 and θ deviating from their "desired values", $\bar{P_1}$ and θ_0. Note that the dependence of P_1 on θ is captured by equation (20) rather than an additional differential equation, which is why the dynamic causal model contains no time-derivative for P_1.

The rest of Beltrami's discussion involves rewriting equations (13), (19), and, (20) such that everything except for θ and ω (including θ_0) gets packed into constants, in order to solve for the equilibrium values of θ, θ' and ω, and then to further explore the system's stability properties. For us, what matters from this part of the discussion is that the equilibrium value of ω inversely depends on the square root of a constant F:

$$F = (P + \alpha cos\theta_0) - \bar{P_1} \tag{21}$$

This dependence confirms that the equilibrium value of ω depends on workload P, though since ω inversely depends on the *square root* of P, this dependence is not very strong.

The reader ought to be puzzled about how to understand ω_0, θ_0, and $\bar{P_1}$. Although Beltrami describes ω_0 as the desired speed, we've seen that equilibrium ω depends on P. Yet we also know that $\theta = \theta_0$ at equilibrium, and that θ's value reflects ω's. Since there are different equilibrium values of ω and θ corresponding to different workloads, and $\theta = \theta_0$ at equilibrium, this entails that we need different values of θ_0 corresponding to different workloads. But θ_0 is treated as an exogenously given constant. Additionally, even if we were to pretend that the governor *could* achieve a fixed equilibrium speed ω_0, the steam required to maintain that speed would not be constant, but would depend on P. So $\bar{P_1}$ ought to depend on P, but is also given as a constant. How does this all fit together?

The answer is that although both θ_0 and $\bar{P_1}$ should depend on P, the effects of changes in P end up cancelling out in equation (21). If, for example P goes up by a certain amount p, then the required amount of steam goes up by that same amount p . And, by (20), the change in $\alpha cos\theta$ equals the $P_1 - \bar{P_1} = p$. So both $\alpha cos\theta_0$ and $\bar{P_1}$ go up by p, and the latter is subtracted by the former. F can therefore maintain a constant relationship to the workload, despite the desired angle and steam depending on the workload. So no harm is done by representing θ_0 and $\bar{P_1}$ as constants. But for the sake of understanding the behavior of the system, it is important to realize that the quantities they supposedly represent are *not* constants, but depend on P.

Pontryagin (1962) explicates θ_0 as a "certain 'mean' value of θ near which the regulated value must be maintained" (p. 217). This acknowledges that θ and ω vary across workloads, without much clarifying things. There is no single value of θ at equilibrium, and treating the constant θ_0 as a mean value does not change this. As we've seen, θ_0 and ω_0 are modeling conveniences, and should not be understood as set points toward which the system aims. They are there to capture the mechanism by which P_1 responds to P, and from the perspective of this mechanism it is incidental that engine speed does not vary by much.

We see that understanding the relationship between P and P_1 is not just a matter of reading the fine print. Even in textbook presentations of the mathematics, the parameters of the equations are defined to suggest that ω is constant across different values of P, without ever explaining why this is OK. Of course, the whole maneuver of distinguishing $\bar{P_1}$ from P_1 is a way of indirectly capturing the influence of changes in workload on steam. But this is all presented in a way to emphasize the derivation of the equilibrium values of θ and ω, and the repackaging of P and P_1 using new parameters for this purpose only further obscures their significance.

The representation of the governor's equations using a dynamic causal model is itself a relatively conservative change. The DCM is better understood as revealing the causal structure latent in the differential equations than as adding new content. The DCM – and its equilibrated counterpart – nevertheless make salient important features of the governor that have been obscured in previous discussions. Given how widely the governor has been studied, this by itself is impressive, and bodes well for applications of DCMs to more complicated systems.

7 Homeostasis, Allostasis, and Cognition

In his 1865 physiology lectures, Claude Bernard kicked off the study of homeostasis by proclaiming: "the constancy of the internal environment is the precondition for free and independent life" (Bernard, 1865, 1974). Discussions of homeostasis that reference control feedback mechanisms often appeal to the notion of a "set point" – e.g. the target temperature to which a thermostat is set – in understanding this constancy. This standard picture of homeostasis may be supplemented and/or contrasted with that of *allo*stasis, which emphasizes not constancy, but rather the longer-term mechanisms by which an organism balances its energy needs (Sterling, 2004). To start clarifying this notion, observe that quantities such as blood pressure or temperature that are allegedly kept constant by homeostasis are not in fact constant, even over the course of a day. These vary between waking and sleeping, and in response to physical activity, and are always varying by at least a bit. On the allostatic picture, the processes regulating these quantities should not be viewed as varying around a set point, but rather as trying to adjust the quantities to match anticipated energy demands in a way that coordinates organisms' competing energy needs across subsystems.

Bechtel and Bich (2024) clarify that Bernard and Cannon did not have the limited view of homeostasis criticized by those advancing modifications or alternatives (such as allostasis). Cannon was aware that homeostasis does not keep a quantity constant, but only within a range, and purposefully used the prefix "homeo" (similar) instead of "homo" (same) when coining the term (Billman, 2020). Nevertheless, the term allostasis is helpful for organizing the study of several significant but neglected features of organismic regulatory systems. Instead of focusing just on the way the body responds to perturbations in order to maintain equilibrium, the allostatic approach focuses on how the body anticipates changes. This is reflected both in terms of its sensory apparatus (e.g. eyes) adjusting to predict the anticipated range of inputs, and the way that certain metabolic processes activate before they are needed (e.g. insulin increasing prior to a meal). The approach uncovers the way different subsystems interact such that one system can borrow energy from another during period of high demand. While the emphasis on homeostatic feedback loops suggests an automatic adjustment process, allostatic mechanisms are connected to the central nervous system, which is needed for balancing incommensurable tradeoffs. Finally, the allostatic approach takes a wider-scoped lens both on biological malfunction and what it takes to repair it. For instance, if an individual has hypertension in response to long-term stress inducing them to be constantly vigilant, a "homeostatic" approach to stabilize blood pressure might fail due to the the body counteracting the stress in other ways. In contrast, the allostatic approach would not localize the problem in any specific response, but rather see it as resulting from a long-term and unsustainable stress reaction that at some point stops being responsive to stress in a fine-grained way (i.e. the vigilance will continue even once external stressors are diminished).

On the standard picture of the Watt governor as aiming to maintain constancy, it is a model for homeostasis (traditionally conceived), and allostasis calls for a new paradigm. We've seen, however, that the governor already performs functions that resemble allostasis. Its function of matching energy supply to demand is more like allostasis than the constancy-maintaining role usually attributed to homeostasis. If the governor were simply a device for maintaining engine speed constant via a feedback loop, an additional mechanism would be required to play the matching role of allostasis. But the governor's causal structure is already capable of matching steam supply to steam demand, and it's equilibrium model highlights this higher-level relationship. The feedback loop is not just a device for keeping engine speed stable, but is also something that can be exploited to create a new relationship between workload and steam flow. By showing how the governor serves to render one variable responsive to another, the DCM for the governor already supplies a key component of allostasis.

A physiologist who (strangely enough) had never thought about 19^{th} steam engine regulators, might be introduced to control systems by being told that homeostatic regulation is like the setting of a thermostat. This brings to mind a set point that is the ideal value towards which one wants room temperature to tend – in applying the metaphor, one would be unlikely to think about the longer-term biological and economic factors influencing which temperature is "ideal". Standard discussions of the governor encourage one to think of it as akin to a thermostat, albeit one where the sensor (which reads temperature) and activator (which turns the heat or air conditioning on or off) are combined into a single device (θ). In contrast, the discussion of the governor here moves away from thinking of the average values of a regulated subsystem as ideal set points towards which that system tends. Rather, it emphasizes that the only sense in which certain values might be "ideal" are relative to the tradeoffs that organisms make to increase their fitness. As Sterling (2004) notes, while proper human functioning requires that certain quantities – e.g., oxygen levels in the brain – remain permanently high, this reflects an evolutionary design choice rather than some more general principle favoring constancy.

The governor does not model all elements of allostasis, as it does not make predictions in any obvious sense. Here it would be premature to speculate on precisely how the models will change as they are developed for increasingly complex systems. But before getting more ambitious, it is crucial to understand of how simpler systems perform their functions, and we've seen that there is more to the governor than meets the eye. In this respect, the lessons drawn here for the governor resemble those drawn by Weinberger and Allen (2022) in the context of cognitive science. Whereas van Gelder (1995) described the governor as a device in which everything influences everything else at the same timescale, with no internal spatiotemporal structure, Weinberger and Allen highlighted the time-scale separation already built into the governor's equations. Their point wasn't that such structure is sufficient for cognitive sophistication, but rather that the mathematical structure present in more cognitively sophisticated systems is already present in the equations for the governor. As one considers more complex cognitive agents, the same basic mathematical devices are employed, and what makes them more complex cannot be read off of generic mathematical features such as feedback loops and differential equations, but rather scales with the complexity of the functions agents perform when adapting to a wider range of environments.

The distinction between integrated and proportional control has potential implications for debates over mental representations. Taking proportional controllers like the governor as models for cognition, it's clear that the governor provides no basis for a representation of the workload-independent speed towards which the governor tends – there *is* no such speed. Yet integrated controllers do need a way to keep track of the system's deviation from a desired speed, and they presumably have different causal structures. So it is open for someone to argue that (the cognitive analogues of) such systems can better be understood as representing the target the system aims at, though here I won't pursue this further. My point is simply that the distinct causal structures of different control systems potentially matter for debates over cognition. Even if one adopts the stance that whether a system has representations is not intrinsic to the system, but rather an assumption adopted by observers when making predictions about it (Dennett, 1989), one may still ask which causal structures are more fruitfully understood by reference to mental representations.

One reason why modeling the governor using Iwasaki and Simon's framework is so fruitful is that it reveals the relationships between causal representations of a system at different scales. The governor's matching function involves the exploitation of a lower-scale feedback loop to bring about a higher-scale relationship, which in turn could be exploited at even higher scales. The strategy of exploiting lower-scale relations for causal control is a very general one for processes evolving towards complexity. This has been studied extensively by Michael Levin and collaborators (Levin, 2023; Seifert et al., 2024), who regularly appeal to the causal structure of goal-oriented systems, without (to my knowledge) referencing any causal models of such structures. His approach is heavily inspired by cybernetics, so the causal modeling of control systems begun here would potentially help fill in the causal details.

8 The Engineering Approach to Causal Modeling

Earlier I noted that interventions are sometimes represented as breaking the arrows into the intervened upon variable. This corresponds to rendering the target variable independent of its causes other than the intervention. Intervening on a variable in an equilibrium model in this way corresponds to a "clamp" intervention, which holds a variable's value fixed indefinitely through time. Weinberger (2021) clarifies how certain equilibrium models apply only under conditions where one does not intervene on certain variables in the equilibrium model, but instead "lets go" and does not intervene upon them. My discussion of the governor furthers the idea that certain causal models only apply under conditions where one does not intervene on particular variables. A clamp intervention holding the value of ω indefinitely fixed would be neither desirable nor compatible with the operation of the governor. The type of control achieved by the governor derives not from fixing ω, but from allowing it to adjust in the way it needs to in order for steam flow to quickly respond to changes in workload.

Although intervention and control are often treated as synonymous, dynamical causal control requires one *not* to intervene on certain variables. This is not just a claim about how we should use these terms, but about how to understand causal control. In modeling control, we need to pay attention not just to regulated variables such as ω, but also to the higher-scale relationship by which P causes P_1 by exploiting the lower-scale feedback loop. Higher-scale causal control is not about keeping lower-scale conditions constant, but rather enabling them to adjust as necessary to yield the higher-scale relationship. Control does not call for formalizing a new type of intervention on a causal model, but involves attending to the lower-scale conditions supporting the relationships in higher-scale models.

Here I do not pursue a general analysis of causal control, but it is worth emphasizing how different such a project would be from standard causal inference. In standard causal inference, interventions are used to test the properties of preexisting structures, and nothing in this literature provides guidance about how to construct systems with causal relationships that would support novel interventions. Engineering of course does allow one to design machines that increase the range of possible interventions,[5] and the dynamic causal modeling framework discussed here provides a bridge between standard causal models and the applied mathematics of engineering. Engineering by its nature involves a fair degree of tinkering. In deciding whether to model a device using the equations for a simple harmonic oscillator (i.e. a pendulum), one presumably does not search over the space of possible models, but starts by trying out a simple model, seeing how it works, and adjusting as needed. Why this should work as well does is a substantive philosophical question, but the fact that it often does work is not in question. An *engineering approach to causal modeling* would involve taking causal models such as the governor's and seeing whether similar models help account for systems found in the world. Undoubtedly, something like this is already going on in more computational analyses in biology and neuroscience, without drawing any systematic connections to causal methods.

The difference between intervention and causal control coheres with the allostatic insight that treatment is not necessarily about restoring a quantity to a desired value, but rather repairing the mechanism so the quantity can adjust as needed. It also has broader implications for understanding when interventions will or won't succeed. Consider the phenomenon of *induced demand* from urban planning (Vanderbilt, 2009). Apparently, one cannot reduce congestion on a busy highway by adding a new lane. Why? Because widening the highway decreases the cost of driving on it, increasing demand until it is congested again.[6] So a highway-widening intervention designed to reduce long-term congestion won't work, which is why urban planners prefer measures that reduce demand, such as congestion pricing. The highway-expanding intervention involves a misunderstanding of the highway's role in the system, which is to meet the demand for people wanting to get from A to B. A highway that is empty most of the time is an inefficient use of resources. Optimal solutions don't avoid having the highway match demand, but rather change people's demand for travel (e.g. by internalizing externalized social costs). We see that the highway's matching function has implications for which variables one should target via interventions.

9 Conclusion

This paper bridges the domains of biology, engineering, and causal inference, yielding payoffs for all three. Within biology, the discussion supports the view that homeostasis does not maintain quantities at fixed values, but rather coordinates the energy needs of distinct organismic subsystems. While this role might appear to call for mathematical tools beyond those supplied by feedback control models, this only seems so because standard treatments of feedback neglect important features that dynamic causal models make salient. Additionally, causally modeling control feedback reveals that, contrary to the common identification of control with interventions, intervention and control can be mutually exclusive. Finally, the link between causation and engineering points towards a novel form of causal inference involving constructing causal systems, rather than merely intervening on found ones.

The discussion here makes salient that higher-scale stability is ensured not by lower-scale constancy, but by designing systems to appropriately adjust to change. This lesson is especially important in the causal context, where the emphasis on interventions skews the discussion towards cases in which variables are held fixed. We are only now beginning to understand how to intervene (or not) to utilize systems' multiscale dynamics. The results here are encouraging, insofar as they suggest in pursuing this project, there is much to learn from unifying the results from independent mathematical domains.

[5] Gopnik (2024) appreciates this point, noting that "designing a gas pedal imposes a relation of mutual information between the pedal and the car acceleration that did not exist before the car was engineered"

[6] This is an instance of "perfect adaptation" (Blom and Mooij, 2023)

Acknowledgements

This paper began as a revision of my unpublished "Homeostasis and the Faithless Foundations of Causal Inference", which benefited from helpful feedback from Colin Allen, Mel Andrews and Nic Fillion, and audiences in Bergen and in Munich, at Lauren Ross' Causation and Explanation workshop. The current version received helpful comments from Ian Shrier, Julia Seidel, and Russell Steele, and benefited from discussions with audiences at Duke and Johns Hopkins. John Dougherty allowed me to repeatedly storm into his office to talk about differential equations, and William Austin Casey provided further helpful discussion. I am extremely grateful to these friends and colleagues for their generous feedback and support, and would further like to thank the Deutsche Forschungsgemeinschaft for the grant that made this research possible (DFG, project no. 504210582).

References

Albert, D. Z. (2003). *Time and chance*. Harvard University Press.

Ball, R. (2003). Control, stability, and bifurcations of complex dynamical systems. *Preprint, Australian National University Centre for Complex Systems*.

Bechtel, W. and L. Bich (2024). Situating homeostasis in organisms: maintaining organization through time. *The Journal of Physiology 602*(22), 6003–6020.

Beltrami, E. (1987). *Mathematics for Dynamic Modeling*. Academic Press.

Bernard, C. (1865). *Introduction à l'étude de la médecine expérimentale par m. Claude Bernard*. Baillière.

Bernard, C. (1974). *Lectures on the phenomena of life common to animals and plants*. Number 900. Charles C. Thomas.

Billman, G. E. (2020). Homeostasis: the underappreciated and far too often ignored central organizing principle of physiology. *Frontiers in physiology 11*, 200.

Blom, T. and J. M. Mooij (2023). Causality and independence in perfectly adapted dynamical systems. *Journal of Causal Inference 11*(1), 20210005.

Blom, T., M. M. Van Diepen, and J. M. Mooij (2021). Conditional independences and causal relations implied by sets of equations. *Journal of Machine Learning Research 22*(178), 1–62.

Boeken, P. and J. M. Mooij (2024). Dynamic structural causal models. *arXiv preprint arXiv:2406.01161*.

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological reviews 9*(3), 399–431.

Dash, D. (2003). *Caveats for Causal Reasoning with Equilibrium Models*. Ph. D. thesis, University of Pittsburgh.

Dash, D. and M. Druzdzel (2001). Caveats for causal reasoning with equilibrium models. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pp. 192–203. Springer.

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Druzdzel, M. J. and H. Van Leijen (2001). Causal reversibility in bayesian networks. *Journal of Experimental & Theoretical Artificial Intelligence 13*(1), 45–62.

Farsinezhad, S. (2015). *Applications of the Kalman filter to the estimation of structural nested mean model*. McGill University (Canada).

Franklin, G. F., J. D. Powell, A. Emami-Naeini, and J. D. Powell (2002). *Feedback control of dynamic systems*, Volume 4. Prentice hall Upper Saddle River.

Gopnik, A. (2024). Empowerment gain as causal learning, causal learning as empowerment gain: A bridge between bayesian causal hypothesis testing and reinforcement learning. presented at the 29th meeting of the philosophy of science association, new orleans, dec 2024.

Hausman, D. M. (1998). *Causal asymmetries.* Cambridge University Press.

Ismael, J. (2023). Reflections on the asymmetry of causation. *Interface Focus 13*(3), 20220081.

Iwasaki, Y. and H. A. Simon (1994). Causality and model abstraction. *Artificial intelligence 67*(1), 143–194.

Janzing, D. and S. H. G. Mejia (2024). A phenomenological account for causality in terms of elementary actions. *Journal of Causal Inference 12*(1), 20220076.

Kang, C.-G. (2016). Origin of stability analysis:\" on governors\" by jc maxwell [historical perspectives]. *IEEE Control Systems Magazine 36*(5), 77–88.

Levin, M. (2023). Darwin's agential materials: evolutionary implications of multiscale competency in developmental biology. *Cellular and Molecular Life Sciences 80*(6), 142.

Maxwell, J. C. (1868). I. on governors. *Proceedings of the Royal Society of London* (16), 270–283.

Mooij, J. M., D. Janzing, and B. Schölkopf (2013). From ordinary differential equations to structural causal models: the deterministic case. *arXiv preprint arXiv:1304.7920*.

Munis, J. R. (2012). *Just enough physiology.* OUP USA.

Pearl, J. (2009). *Causality.* Cambridge university press.

Pontryagin, L. S. (1962). *Ordinary Differential Equations: Adiwes International Series in Mathematics.* Addison-Wesley Publishing Company Inc.

Rubenstein, P. K., S. Bongers, B. Schölkopf, and J. M. Mooij (2016). From deterministic odes to dynamic structural causal models. *arXiv preprint arXiv:1608.08028*.

Seifert, G., A. Sealander, S. Marzen, and M. Levin (2024). From reinforcement learning to agency: Frameworks for understanding basal cognition. *Biosystems 235*, 105107.

Simon, H. (1953). Causal ordering and identifiability. In W. Hood and T. Koopmans (Eds.), *Studies in Econometric Method. New York: John Wiley and Sons*, pp. 49–74. New York: John Wiley and Sons.

Simon, H. A. and N. Rescher (1966). Cause and counterfactual. *Philosophy of science 33*(4), 323–340.

Spirtes, P., C. Glymour, R. Scheines, et al. (2001). Causation, prediction, and search. *MIT Press Books 1.*

Sterling, P. (2004). Principles of allostasis: optimal design, predictive regulation, pathophysiology, and rational. *Allostasis, homeostasis, and the costs of physiological adaptation 17.*

Sterling, P. (2020). *What is health?: Allostasis and the evolution of human design.* MIT Press.

Strogatz, S. H. (2024). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* Chapman and Hall/CRC.

van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy 92*(7), 345–381.

Vanderbilt, T. (2009). *Traffic: Why we drive the way we do (and what it says about us).* Vintage.

Vyshnegradsky, I. (1876). On the general theory of governors (sur la théorie génerale des régulateurs).

Weinberger, N. (2019). Reintroducing dynamics into static causal model. In S. Kleinberg (Ed.), *Time and Causality Across the Sciences.* Cambridge: Cambridge University Press.

Weinberger, N. (2020). Near-decomposability and the time-scale relativity of causal representations. *Philosophy of Science*.

Weinberger, N. (2021). Intervening and letting go: On the adequacy of equilibrium causal models. *Erkenntnis*, 1–25.

Weinberger, N. and C. Allen (2022). Static-dynamic hybridity in dynamical models of cognition. *Philosophy of Science 89*(2), 283–301.

Wiener, N. (2019). *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press.

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

Zhang, J. (2022). On the unity between observational and experimental causal discovery. *Theoria: An International Journal for Theory, History and Foundations of Science 37*(1), 63–74.