Dynamical Causes

Russell Meyer*

**School of Humanities & Social Inquiry, University of Wollongong, Wollongong, NSW, Australia.*

Corresponding Author:

Mr. Russell Meyer

e-mail: rjm322@uowmail.edu.au

ORCID: 0000-0003-3088-2617

Keywords:

Mechanistic Explanation, Mechanism, Dynamical Models, Dynamical Explanation, Cell Fates, Systems Biology

Acknowledgements:

I am particularly grateful to Nick Brancazio for minor comments and major support. Many thanks also to Michael Kirchhoff for helpful comments on this manuscript.

Dynamical Causes

Abstract

Mechanistic explanations are often said to explain because they reveal the causal structure of the world. Conversely, dynamical models supposedly lack explanatory power because they do not describe causal structure. The only way for dynamical models to produce causal explanations is via the 3M criterion: the model must be mapped onto a mechanism. This framing of the situation has become the received view around the viability of dynamical explanation. In this paper, I argue against this position and show that dynamical models can themselves reveal causal structure and consequently produce non-mechanistic, dynamical explanations. Taking the example of cell fates from systems biology, I show how dynamical models, and specifically the attractor landscapes they describe, identify the causes of cell differentiation and explain why cells select particular fates. These dynamical features of the system better fit Woodward's (2010, 2018) criteria of specificity and proportionality and make them the best candidate causes of cell fates than mechanisms. I also show how these causes are irreducible and inaccessible to mechanistic models, making 3M unworkable and counterproductive in this case. Dynamical models can reveal dynamical causes and thereby provide causal explanations.

1. Introduction

The concept of causal structure of the world, or just *causal structure,* is a touchstone of modern Mechanism[1] originating from Salmon's (1984) causal-mechanical approach to scientific explanation. The claim that a close relationship exists between causal structure and explaining a phenomenon is outlined by Craver (2007):

> "There are perhaps many interesting things to be said about explanatory texts, but one crucial aspect of their adequacy has to do with whether explanatory texts accurately characterize the causal structure of the world." (Craver 2007, pg. 27)

A related and complementary claim made by mechanists is that scientific explanations should describe the causal relationships which comprise this causal structure:

> "In many areas of science, explanations are said to be adequate to the extent, and only to the extent, that they describe the causal mechanisms that maintain, produce, or underlie the phenomenon to be explained, the explanandum phenomenon." (Kaplan & Craver 2011, pg. 601).

Putting these ideas together, I take causal structure to refer to the comprehensive web of causal relations that underlie or produce a phenomenon. Something explanatory ought to follow on from having a description of causal structure – if you understand all the relationships driving a phenomenon to occur, you have explained it.[2] In short, explanation is all about describing causal structure.

---

[1] Following the convention proposed by Glennan & Illari (2018b) I distinguish the philosophical stance of Mechanism from the object called a mechanism via a capitalisation added to the former.

[2] Though Kaplan & Craver (2011) qualify this demand for completeness: descriptions that reveal a partial causal structure and are in the process of completion can also be considered explanatory.

And as mechanists have argued, mechanisms have a special, if not unique, role to play in this revelatory task:

> "Mechanisms explain the diverse aspects of the explanandum phenomenon, and so unify them by relating them to an underlying causal structure" (Craver 2007 pg. 49)

> "[Mechanistic] models…carry explanatory force to the extent, and only to the extent, that they reveal (however dimly) aspects of the causal structure of a mechanism." (Kaplan & Craver 2011, pg. 602)

This closely pairs the notion of causal structure to explanatory power, and mechanisms link the two because, in many cases, a description of a mechanism is necessary for describing causal structure. Driving the point home, these mechanists argue that explanations necessarily capture the totality of causal relations producing a phenomenon, and that this totality – the causal structure – is in large part what differentiates a loose bundle of descriptions from a genuine explanation (Craver 2006, 2007).

But not everyone is satisfied with this mechanistic arrangement. A movement to articulate a mode of non-mechanistic[3], dynamical explanation based on dynamical models has been afoot for some time (Chemero & Silberstein 2008, Stepp et al. 2011). Dynamical models are a kind of mathematical model that employs the tools of dynamical systems theory to capture the unfolding of variables over time using differential and difference equations. These models have a track record of impressive descriptive accuracy and predictive power when applied to various cognitive, neuroscientific and

[3] A nonmechanistic explanation refers broadly to any explanation that does not appeal to underlying causal mechanisms for its explanatory power.

biological phenomena. By their very nature they do not by themselves give much detail about the physical realisers or substrates of the variables they model: "[s]uch dynamical systems explanations…don't seem to proceed by the decomposition into parts with intrinsic behaviours that is characteristic of modular styles of explanation." (Woodward 2013, pg. 60). Proponents consider this advantageous – because of this feature dynamical models can "zoom out" from these fine-grained details and say new and interesting things about the dynamical features of a system in a way mechanistic models do not, since even the most minimal mechanistic account is committed to describing the concrete entities and activities producing phenomena (Glennan & Illari 2018a).

Breaking with mechanist views on explanation, proponents of dynamicism reject the necessity for mechanistic models as a prerequisite for explanation and emphasise accuracy in description and prediction as sufficient (Chemero & Silberstein 2008). Dynamicists have mostly conceded that dynamical explanations need not adhere to mechanist standards around explanation, specifically the requirement that they ought to describe causal mechanisms (Stepp et al 2011). The dynamicist account further departs from the mechanist orthodoxy by suggesting that dynamical explanations "need not respect the underlying causal structures that give rise to system-level dynamics." (Kaplan & Craver 2011, pg. 602).

Mechanist philosophers have, as a result, been critical of claims that dynamical models are explanatory. A more thorough justification for Chemero & Silberstein's (2008) claim that prediction and counterfactual description are sufficient for explanation is missing – and as Kaplan & Craver (2011) point out, much more work would be required on the dynamicist's part to demonstrate that predictive power yields explanatory power.

Further, if mechanists are right that causal structure is core to explanation, and mechanisms are crucial to getting at causal structure, then on two counts dynamical models are a non-starter as standalone explanations. This general concern, termed the *causal relevance concern* (Meyer 2018) is the biggest hurdle for getting dynamical explanation up and running.

According to mechanists, the solution to the causal relevance concern is straightforward: associating models with a mechanism in line with the model-to-mechanism-mapping (3M) requirement (Kaplan 2015; Kaplan & Craver 2011). The idea behind 3M is that so long as the terms in a dynamical model can be associated with (mapped onto) the mechanistic components underlying the model, then the dynamical model can thereby describe causes:

> "(3M) A model of a target phenomenon explains that phenomenon to the extent that (a) the variables in the model correspond to identifiable components, activities, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these (perhaps mathematical) variables in the model correspond to causal relations among the components of the target mechanism." (Kaplan 2011, pg. 347).

By being grafted onto a mechanism, dynamical models do say something about the causal structure underlying the phenomenon, namely the temporal and organisational features of the causal relations between mechanistic components, a view also developed and endorsed by Bechtel & Abrahamsen (2010, 2013). Similarly, when Craver & Kaplan (2018) claim that "[n]ot all dynamical models describe causal relations. Explanatory dynamical models do…" it means that properly mechanistic models (with added

dynamical details) are explanatory. The status of dynamical models, according to mechanists, is therefore as descriptive tools in service to mechanistic explanations.

On their own, dynamical models do not describe causal (and hence explanatory) relationships[4], but rather function as a useful tool for describing the temporal organisation of mechanisms (Kaplan 2015). Their utility acknowledged, dynamical models are still in an asymmetrical relationship with mechanisms, "their explanatory value can be seen as clearly depending on the presence of an associated account (however incomplete) of the parts in the mechanism" (Kaplan 2015 pg. 760). There is no causal story a dynamical model can provide that does not, ultimately, boil down to a mechanistic model.

In this paper I will argue against this mechanist interpretation of dynamical models, and provide a novel example of non-mechanistic, dynamical explanation at work. First, I will contest the claim that dynamical models do not describe causal relations, appealing to an even-handed application of interventionist standards (Meyer 2018). This clears the immediate path to allow dynamical models to give descriptions of causal structure. Second, I will show how dynamical models of cell fates, an example borrowed from systems biology, uncover the causal structure underlying this phenomenon. The attractor landscape described by this model, I will argue, reveals the causes of cell differentiation without reference to mechanisms. Thirdly, following from this positive account, I use Woodward's (2010, 2018) and Waters' (2007) notions of specificity and proportionality to further illustrate how attractors are the best candidate difference-makers for cell fate outcomes, and best describe the causal structure of the phenomenon. Finally, I dispute

[4] It should be mentioned here that there are ongoing discussions regarding the feasibility of *non-causal* dynamical explanations (e.g. Ross 2015; Chirimuuta 2017). I will however focus specifically on the case for *causal* dynamical explanations and bracket the non-causal option.

that the causal structure identified by dynamical models of cell fates can or ought to map onto mechanistic models, per the 3M requirement.

2. Causal Structure and Dynamical Models

Recently several arguments have been made targeting the causal relevance concern generated by the foregoing mechanist picture of explanation (Meyer 2018; van Eck 2018). These authors have claimed that dynamical models can in fact describe causal relations and thereby explain non-mechanistically, while retaining the mechanist's own interventionist framework to make their case.

The main strains of Mechanism all appeal to Woodward's (2003) interventionism to support the notion that mechanisms describe causes. This should not be taken to imply that interventionism privileges mechanisms as the only possible source of causal relations. On the interventionist account, mechanistic explanations are *"one* important variety of explanation" capable of "meeting the general interventionist conditions on explanation" (Woodward 2017, pg. 85) among other candidates. On Meyer's (2018) view, if dynamical models can similarly meet these general interventionist conditions for establishing causal relations, then they ought to be considered causal.

What are Woodward's general interventionist conditions? Woodward supplies (M):

> (M) X causes Y if and only if there are background circumstances B such that if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y would change. (Woodward 2008, pg. 222).

(M) specifies how ideal interventions can be used to establish causal relevance of variables. These interventions establish the relationship between the value of a variable, X, and the value of a variable Y. Changes in Y which are the direct result of changes in X

demonstrate a causal relationship. X is causally relevant to Y if (M) is satisfied. Hence if a variable in a dynamical model can meet the requirements of (M), then it ought to be considered a cause.

Meyer (2018) uses the example of the Haken-Kelso-Bunz (HKB) (Haken et al. 1985) model of bimanual coordination to demonstrate how dynamical models can describe causes. Bimanual coordination is the phenomenon whereby synchronised movements on either hand (in this case moving index fingers from side to side) can be coordinated to move either in-phase, or anti-phase (where ϕ=0 and 180 respectively). To accomplish this, the HKB model uses a differential equation to map the system's evolution over time:

$$\frac{d\phi}{dt} = -a \sin \phi - 2b \sin 2\,\phi$$

Where ϕ represents relative phase, ranging between 0 degrees and 180 degrees (in- and anti-phase conditions respectively); and b/a relates the frequency of oscillations.

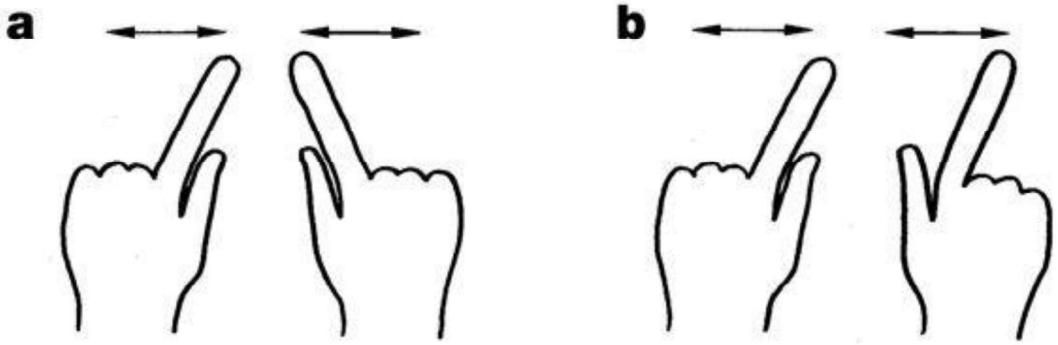

Figure 1: An illustration of bimanual coordination, where (a) represents in-phase coordination, and (b) represents anti-phase coordination. Reproduced from Mechsner et al (2001).

In order for the relation from b/a and ϕ to be causal, the following would need to hold:

> "(M) b/a causes ϕ if and only if there are background circumstances B such that if some (single) intervention that changes the value of b/a (and no other variable) were to occur in B, then ϕ would change." (Meyer 2018, pg. 13)

Experimental interventions into this system involve changing the frequency of oscillations (b/a) in order to observe their effect on relative phase (ϕ). Subjects are tasked with attempting to maintain bimanual coordination in either the in- or anti-phase conditions, and also match their frequency of their movements to cues given by the experimenters. This paradigm was used in Scholz & Kelso (1989), who intervened to increase and decrease the frequency of the cues provided to the subjects. Several predictions of the HKB model were validated in the results of these experiments: when oscillation frequency is slow (b/a > 0.25) both in-phase and anti-phase patterns of coordination are quite stable – subjects are able to maintain these movements without altering phase. But at higher frequencies (b/a =< 0.25) the anti-phase pattern (ϕ = 0) becomes difficult to maintain, and subjects tend to slip into an in-phase pattern (ϕ = 180).

The key point here is that the relationship between b/a and ϕ is not merely a correlation. There is a direction established by this experimental intervention from b/a to ϕ, from cause to effect. Further these interventions produce regular, function-like changes in the value of ϕ. The variable b/a is the difference-maker to ϕ – while other variables may provide necessary background conditions (B), it is b/a that causes ϕ to change in value. So in this case, (M) should be satisfied as a straightforward example of a causal relation –

the variables in the HKB model describe causes. In Woodwardian terms, they describe difference-makers, the systematic relationships from cause to effect. [5]

3. Cell Fates

Having outlined the basic framework for causal, dynamical explanation, I now turn to the novel case of *cell fates*, a topic of significant interest in contemporary systems biology and cell genetics. An almost ubiquitous feature of animal cells is their capacity for differentiating into cell fates – alternate, stable phenotypes expressing new traits. Frequently one kind of primogenitor (undifferentiated) cell can differentiate into several distinct cell fates, each exhibiting a different phenotype. Some stem cells, for instance, are bi- or multi-potent, meaning they have the potential to transition into two or more stable phenotypes respectively.

These fates are interesting for a few reasons. Firstly, they tend to be stable and the transitions into them reliably one-directional under normal circumstances. Once differentiated, cells do not tend to "un-differentiate" backwards into progenitor cells or switch over into a different fate. Secondly, cells tend to transition through a series of phenotypes in between the progenitor phenotype and cell fate phenotype in a very

[5] I acknowledge here the significant debates around higher-level interventions in the mechanist literature, particularly the problem of *fat-handedness*: intervening on a higher-level variable necessitates simultaneously intervening on its supervenience base, and hence violating the interventionist requirement for isolating a single variable for intervention (see Baumgartner & Gebharter 2017, Krickel 2017). I bracket this substantial discussion by adhering to Woodward's (2015) clarification to (M). Woodward specifies that non-causal supervenience relations between micro- and macro-levels need not be held steady in the same fashion as causal relations, so that "properties that supervene on but that are not identical with realizing properties can be causally efficacious." (Woodward 2015, pg. 303)

directed fashion – even if external perturbations disrupt this typical course, they still find their way to a stable fate.

One story invoked to explain these features of cell fates is a kind of genetic pre-destination. Each phenotype, on this interpretation, must contain some kind of instruction for how to progress to the next phenotype, and that phenotype to the next, and so on down the line. Alternatively, the cell may receive external signals that help direct and drive these transitions and maintain them.

These conceptions have proven too coarse-grained in many situations (Huang 2012). More recent developments in cell genetics and systems biology suggest that differentiations are driven not in a step-by-step or externally controlled fashion, but in a self-organised process driven by networks of thousands of genes engaged in extremely complex interdependent relationships, with all kinds of endogenous activity determining the transitions between fates, as well as their relative stability.

The need to make sense of these complex relationships is in part responsible for the development of gene regulatory networks (GRNs) as a modelling tool. A GRN is a network map of the relationships between the many genes that make up the genotype of a given cell, a series of "layers of molecular regulatory networks and cell-cell communication networks – a web of interactions through which genomic information must percolate to produce the macroscopic phenotype" (Huang 2012, pg. 153). These "interactions" consist of each gene's expression behaviour, namely what proteins it instructs a cell to transcribe, and how this influences the activities of other genes. How these expressions promote, inhibit, and otherwise interfere with the expressions of other genes is what makes up the architecture of a GRN.

Unsurprisingly given the number of interconnected transcription processes involved, what GRNs help illustrate is that there is no fixed set of genetic tracks that determines a cell's transitions. While some interactions can be identified as important in particular transitions, what decides the cell fate of a given cell is a highly complex, high-dimensional network involving thousands of interconnected genes (Huang et al. 2005). What GRNs can show is which phenotypes are stable or unstable relative to their neighbours, and how these differences can drive transitions to new phenotypes.

GRNs are the first tool used in developing an explanation of cell fates. The second is Waddington's (1957) notion of epigenetic landscapes. Waddington's metaphor has proven particularly durable and appealing to many biologists concerned with cell fate phenomena, and the metaphor appears frequently in this scientific literature (e.g. Enver et al 2009; Davila-Velderrain et al 2015; Moris et al 2016). Adherents to this line of thinking equate progress through stable and unstable phenotypes to progress through the peaks and valleys of epigenetic landscapes, with the phenotype represented by a ball rolling through this terrain, like in Waddington's famous illustration.

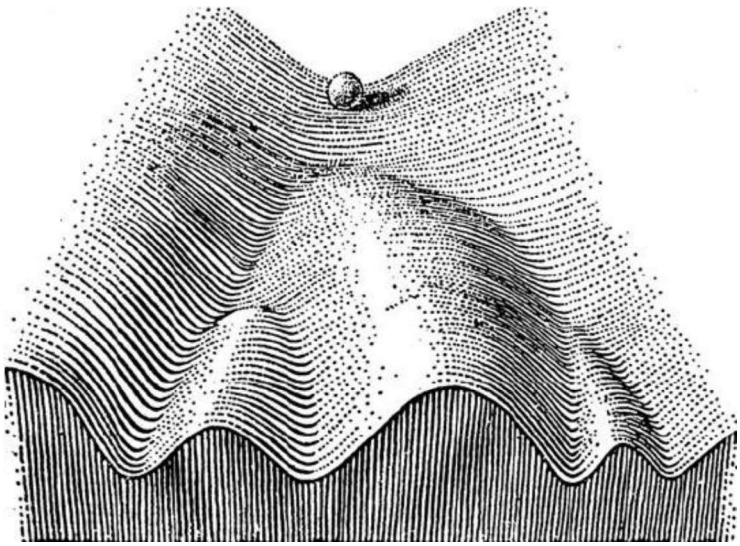

Figure 2: Waddington's illustration of the epigenetic landscape. The ball (phenotype) runs down through the landscape and is canalised into different phenotypic outcomes. From Waddington (1957).

This is where dynamical models enter the picture. In order to put Waddington's ideas about the epigenetic landscape into practice, researchers appeal to dynamical models as a way of capturing the trajectory of a cell through the many possible phenotypes it could express. The resulting models and attractor landscapes bear a striking resemblance to Waddington's landscape, which has as a result has been reconsidered from merely illustrative metaphor to something potentially more revealing about the workings of GRNs (Jaeger & Monk 2014).

Attractor landscapes are a frequently used visualisation of dynamical models, and are virtually ubiquitous in models of cell fates. To produce a 3D model that illustrates a cell's trajectory through different states (phenotypes), modellers reduce the many dimensions (genes) involved, of which there may be thousands, into a plane. Each point on that plane represents a possible state for the system to inhabit, and nearby states represent similar states. The relative height or depth of any given point indicates its stability or lack thereof.

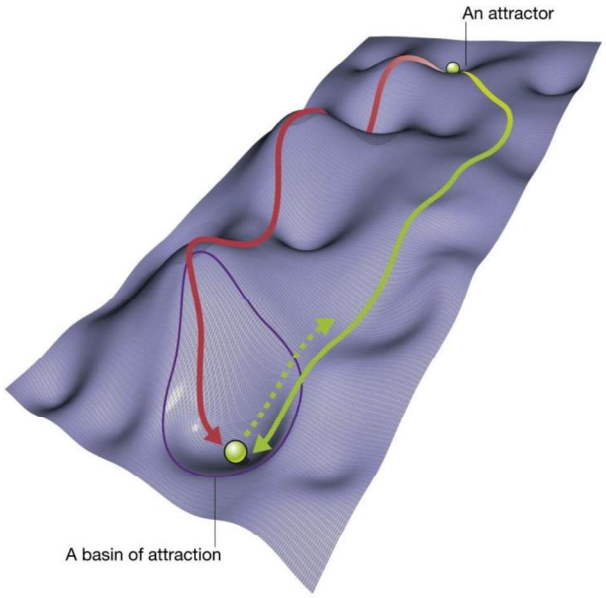

Figure 3: An attractor landscape. The green ball represents the phenotype, which has settled into one of several available basins of attraction. From Enver et al (2009).

In dynamical systems theory an attractor is a point in this landscape that represents a stable solution to the equations that make up the model. Over time, the system will converge towards an attractor if it enters its basin of attraction. A basin of attraction is the set of points that "feed into" a given attractor. The convergence on an attractor might be stable, where the system settles right on the attractor point – or it may oscillate around that point ("circling the drain") for some time or even indefinitely.

In the attractor landscape these features are visualised as the troughs and valleys that Waddington's "ball rolling down the hill" follow and settle into. These attractor landscapes will, later in this section, be shown to provide the causal detail needed to describe and explain cell fates.

Biologically, the attractor point itself corresponds to a stable phenotype – a cell fate.

> "Extrapolating from Waddington, different cell types may be seen as stable solutions of transcription factor networks—or 'attractors'—which occupy the basins of Waddington's landscape." (Graf & Enver 2009, pg. 590)

Figure 4 illustrates a toy example characteristic of many GRNs. There are two proteins being transcribed, *a* and *b*, each of which inhibits the transcription of the other – *mutual inhibition* – and encourage transcription of themselves in a positive feedback loop – *auto-stimulation*. These processes of mutual inhibition and auto-stimulation are common features of cells that are multi-stable.

There are also three attractors present. The first, *a/b*, is the stable starting point – the undifferentiated, progenitor phenotype where neither *a* nor *b* is transcribed at a high rate, and from which the system is unlikely to budge. If left undisturbed, the effects of mutual inhibition and auto-stimulation will generally ensure that transcription of *a* and *b* remains roughly even. Attractors *a* and *b* represent two "downhill" cell fates the perturbed system may end up in – if *a/b* were destabilised, the system will bifurcate, and converge on either *a* or *b*.

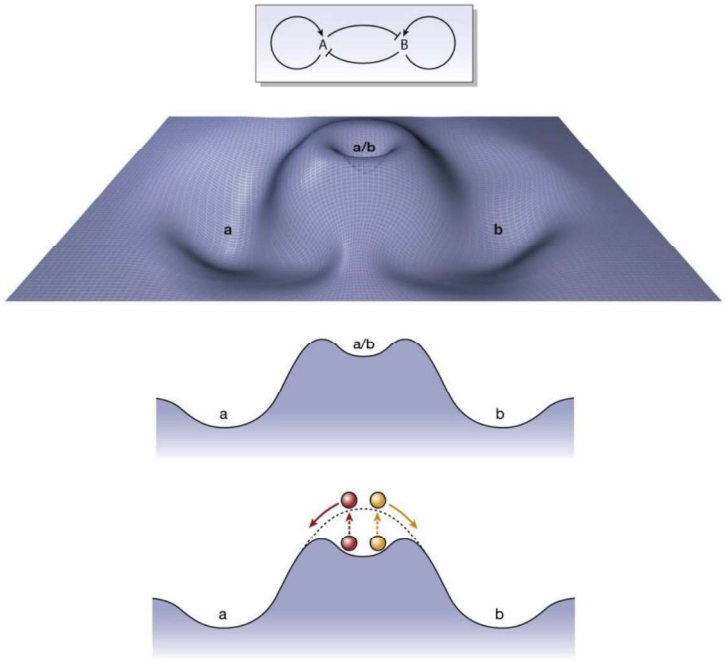

Figure 4: The attractor landscape of the toy cell fate system. The system is depicted bifurcating from a/b, the progenitor fate, into either the a or b fate. From Enver et al (2009).

Graduating from a toy model, I turn now to a real-world example: Huang et al's (2007) model of FDCP-mix cells, a kind of bone-marrow cell. An FDCP-mix cell is capable of differentiating into two distinct cell fates called *erythroids* and *myeloids*. Differentiation into erythroid/myeloid is influenced heavily by two transcription factors, *GATA1* & *PU.1* respectively. GATA1 & PU.1 are both auto-stimulating, and mutually inhibiting. The following model describes the activation and regulation of GATA1 and PU.1:

$$\frac{dx_1}{dt} = a_1 \frac{x_1^n}{\theta_{a_1}^n + x_1^n} + b_1 \frac{\theta_{b_1}^n}{\theta_{b_1}^n + x_2^n} - k_1 x_1$$

$$\frac{dx_2}{dt} = a_2 \frac{x_2^n}{\theta_{a_2}^n + x_2^n} + b_2 \frac{\theta_{b_2}^n}{\theta_{b_2}^n + x_1^n} - k_2 x_2$$

Where x_1 represents GATA1 activity, x_2 represents PU.1 activity, and a_1 & a_2, b_1 & b_2, k_1 & k_2, and θ, all represent control parameters. Parameters a_1/a_2 represent the relative strength of auto-stimulation of GATA1 and PU.1 respectively; b_1/b_2 describe the rate of mutual inhibition of GATA1 and PU.1; k_1/k_2 represent the rate of deactivation of GATA1 and PU.1; and θ represents the strength of the regulatory interaction. Much like the previous toy model, this real-world system exhibits tristability – *C*, the progenitor state, as well as *A* and *B*, the differentiated erythroid and myeloid cell fates.

By intervening on these parameters, Huang et al (2007) were able to alter the features of the attractor landscape. This is analogous to reshaping Waddington's landscape, where a change in the topography of that landscape – which states are stable and unstable relative to their neighbours – can induce a previously stable system to differentiate, and in this instance bifurcate into new "downhill" cell fates, an arrangement referred to as a "transition via a bifurcation" (Huang et al 2007, pg. 701).

Huang et al (2007) focus on the a, b and k parameters since these collectively represent the various regulatory influences on x_1 and x_2. The effects of mutual inhibition, auto-stimulation and deactivation over time of GATA1 and PU.1 are thought to maintain the stability of the system in its progenitor fate, C. Disturbing these variables, then, has the potential to destabilise C and induce differentiation into either A or B.

The model predicts that reducing the rate of auto-stimulation (a_1 & a_2) and deactivation (k_1 & k_2) will destabilise C and induce differentiation. The model also considers an alternative scenario where – due to different initial values in control parameters – the only attractor present in the model is C, the basin of which covers the entire phase space. In this scenario, reducing the value of b_1 & b_2 destabilises C, and also leads to the appearance of the A and B attractors. In either situation, C is converted from a stable state

into an unstable "hill-top", from which any small stochastic variation in GRN activity is enough to induce a differentiation event. The system is compelled to leave C, and converge towards either A or B, the erythroid and myeloid cell fates respectively.

After this initial differentiation event where C can no longer be occupied by the system, the decision to converge on A versus B cell fates becomes available for the cell. Assuming absolute symmetry in the system (A and B are equally accessible from the now destabilised C) then minor variations in initial starting position, and random fluctuations in expression, will be responsible for pushing the system towards A or B. However, asymmetries in the attractor landscape can make one of A or B more accessible than the other and bias the system towards a particular cell fate despite the stochastic nature of the system's trajectory. For instance, a greater value of x_1 versus x_2 will bias the system towards settle into the myeloid state versus the erythroid fate, and vice versa. Huang et al (2007) describe this as "tilting the watershed" in order to "harness and bias the stochastic processes." (pg. 710)

Having considered the phenomenon described by Huang et al's (2007) model (initial differentiation followed by cell fate selection) the question of interest is: how or why do FDCP cells differentiate into erythroid or myeloid cell fates? Put into interventionist speak, we are interested in the *what-if-things-had-been-different question*, or *w-question* about cell fates: counterfactually, under what conditions would we have observed a different result (the selected fate)?

In interpreting this model, I argue that what makes the difference to the outcome, and what answers the w-question, is the attractor landscape. Specifically, I argue the presence or absence of particular attractors is decisive to the outcome for a cell in the process of selecting a fate.

As mentioned earlier, the initial destabilisation event occurs only once C has been intervened upon. If C is present (with sufficiently high ridges to prevent stochastic fluctuations from pushing the cell out of C) then the cell will not undergo a differentiation event. This is ultimately what answers the w-question: a different result would have been obtained depending on the presence or absence of the stable progenitor attractor state C. While the destabilisation of C can be induced via different stimuli (changes to the rate of auto-stimulation, mutual inhibition, and deactivation of GATA1 and PU.1) what makes the difference is C. An investigator intervenes on C through these control parameters. Further, no control parameter stands out as the singular cause of differentiation. In fact, most of the control parameters are, if intervened upon, capable of destabilising the stable attractor C under the right background conditions.

After the initial destabilisation of C, the cell needs to select a new cell fate to differentiate towards from this new unstable position:

> "Metaphorically, the destabilization and disappearance of the progenitor attractor can be viewed as S being placed on a "watershed" region in Waddington's epigenetic landscape (Waddington, 1957) where it can easily be "tipped" into either side to the now easily accessible attractors of the two prospective lineages by small, deterministic perturbations or by random fluctuations in molecular activities, to reliably produce distinct and specific outcomes. This near-symmetric bifurcation model thus is consistent with the ample evidence for the observed stochasticity in fate determination…" (Huang et al 2007, pg. 709-710)

Huang et al (2007) are here comparing Waddington's metaphor with their observed results: a symmetrical destabilisation event leaves the cell equally likely (all other things being equal) to differentiate towards A or B. The general stochastic variation in gene

expression is enough to push the cell towards either one, as well as all manner of incidental factors:

> "…the "watershed" metaphor explains the observation that many unspecific (hence, non-instructive) signals, such as solvents or mechanical forces, can cause differentiation in many cell systems: they may do so by "tipping" cells into a predefined program" (Huang et al 2007, pg. 710)

From a position of high instability, it takes very little, up to and including incidental mechanical forces acting on the cell, to trigger differentiation. Small variations in initial conditions also have an influence, according to the model, on whether A or B is ultimately selected.

But what makes the difference outside this inherent stochasticity in the system is, as mentioned earlier, the biasing of these processes by shaping the attractor landscape. By intervening on control parameters, the investigators were able to "tilt the watershed", making A or B occupy more or less of the phase space with their basin of attraction. This is achieved through an asymmetrical intervention on the control parameters such that x_1 is greater than x_2 (or vice versa). This effectively makes one cell fate – myeloid or erythroid – more accessible from the cell's present state than the other.

The takeaway from this part of Huang et al's (2007) discussion is that there are many events than can induce the initial differentiation event (the destabilisation of C) and many events than can determine whether A or B is selected subsequently, ranging from the various control parameters to external forces. Whatever stimulus is involved, what makes the difference to the outcome is the specifics of the attractor landscape. How accessible a fate is from the cell's present state is what makes the outcome more likely

amidst a barrage of stochastic fluctuations, and what makes a fate accessible are the dimensions of the associated attractor.

4. Dynamical Causes

An objection to the foregoing interpretation would be to question the role of attractors (and really any dynamical feature of a system) as the cause of differentiation in favour of a mechanistic interpretation. If we assume the initial destabilisation event has occurred in a cell, and the system is poised to differentiate to a myeloid or erythroid, perhaps differentiation can simply be explained by the levels of different transcription factors present. For instance, in Huang et al's model, one might prefer a mechanistic interpretation wherein intervening on levels of GATA1 and PU.1 in a cell leads to different outcomes. Higher levels of GATA1 and PU.1 in a cell do indeed predict a higher likelihood of those cells differentiating into erythroids and myeloids respectively (Huang et al 2007). On this reading it seems like the cause of differentiation is decidedly mechanistic. If this were the case, it would of course be fatal to the account being built here.

To air out this criticism, I will consider two possible readings: the mechanist one spelled out above, and the dynamical interpretation I advanced in the previous section, and place them into Woodward's (M) criterion:

> (M1) *GATA1 transcription level* causes *erythroid differentiation* if and only if there are background circumstances *B* such that if some (single) intervention that changes the value of *GATA1 transcription level* (and no other variable) were to occur in *B*, then *erythroid differentiation* would change.
>
> (M2) *Presence of erythroid attractor* causes *erythroid differentiation* if and only if there are background circumstances *B* such that if some (single) intervention that

changes the value of *presence of erythroid attractor* (and no other variable) were to occur in *B* then *erythroid differentiation* would change.

(M1) represents a 3M-compliant interpretation, where the causal relations in the dynamical model map to the underlying mechanistic model. (M2) advances the dynamicist argument – that the causal relations identified by the dynamical model are indeed genuine causes. Selecting which of these scenarios is the better interpretation requires some grappling with how exactly we might select between different potential causes.

Fortunately, Woodward (2010, 2018) has elaborated some criteria designed to clarify situations where the role of difference-maker is ambiguous. The ambiguity arises because not all potential causes are created equal – they "…can differ in the extent to which they satisfy other conditions relevant to their use in explanatory theorizing" (Woodward 2018 pg. 1). Two of these conditions developed by Woodward are I think particularly relevant to the current discussion – these are *specificity*, and *proportionality*.

Taking these conditions one at a time, I want to consider (M1) and (M2) first in terms of specificity. Specificity refers to "a kind of fine-grained and specific control" (Woodward 2010 pg. 306) that a cause has over the outcomes for some effect variable. To motivate the importance of this condition, Woodward draws on an example offered by Waters (2007) who discusses the synthesis of RNA molecules by DNA.

In this case, DNA provides genetic information to RNA polymerase, which then produce RNA molecules. The question is over whether DNA or the RNA polymerase is the cause of RNA molecule output:

> "DNA is a specific difference maker in the sense that different changes in the sequence of nucleotides in DNA would change the linear sequence in RNA molecules in many different and very specific ways. RNA polymerase does not have this specificity…it is not the case that many different kinds of interventions on RNA polymerase would change the linear sequence in RNA molecules in many different and very specific ways.
>
> This shows that DNA is a causally specific potential difference maker. The fact that many such differences in DNA do actually exist and these differences actually explain the specific differences among RNA molecules indicates that DNA is the causally specific actual difference maker..." (Waters 2007, pp. 574–575)

Of the two possible causes, RNA polymerase is the least specific. It does not provide a fine-grained description of the dependency between different states of the cause, and different states of the effect. Either the polymerase is doing its job, or it is not. This is comparatively coarse-grained when we want to understand the specifics of these dependencies.

On the other hand, DNA seems a far more satisfactory candidate cause. There is a fine-grained dependency between the different states of the DNA (what genetic information it inputs) and what RNA molecule is produced. Hence DNA is more specific and is the better candidate for the genuine cause of RNA molecule output.

Specificity requires that states of a cause map uniquely to states of an effect. The less "overlap" or lack of uniqueness in these mappings, the better. This mapping should describe exploitable counterfactual relationships between cause and effect in this fine-grained way. When selecting between candidate causes, especially in biology, the more

specific a cause-effect mapping the more we should be inclined to select it as the genuine cause.

With specificity in mind, let us consider the viability of (M1). Does transcription of GATA1 exercise a fine grained and specific control over the outcome of cell differentiation? GATA1 does increase the probability that a destabilised cell will end up as an erythroid, so it exercises that much control. But that relationship does not tell us why a certain threshold of perturbation is required to initiate differentiation. Nor does it tell us why certain states of the system will differentiate into erythroids, and why others won't. Consequently, there is a considerable amount of overlap between many different values of differentiation and the consequent states of the cell fate. Consequently, the description here is fairly coarse-grained and non-specific.

On the contrary if we accept (M2), then we are delivered far more explanatorily relevant, fine-grained dependencies. The threshold required to initiate differentiation into an erythroid is now explained – it is due to specific stabilities of phenotypes compared to their neighbours, as described by the dynamical model. The fact that some states converge on the erythroid fate and not others is due to the dimensions of the basin of attraction corresponding to that fate.

The second condition to be examined, proportionality, "has to do with the extent to which a causal claim fully captures conditions under which variations in some phenomenon of interest occur." (Woodward 2018, pg. 1). Woodward provides an illustrative example: imagine I train a pigeon to peck at a red stimulus via classical conditioning. I present a new stimulus to the pigeon, and it pecks at it. Two possible causal claims can be introduced to describe what has just happened:

1. The presence of a scarlet stimulus caused the pigeon to peck.

2. The presence of a red stimulus caused the pigeon to peck.

1 is not untrue. Scarlet is a type of red, and on this basis, one could fairly claim that a scarlet stimulus caused the pigeon to peck. However there seems to be something off about this interpretation. Woodward identifies the flaw in that the "scarletness" of the stimulus is not what induced the pigeon's pecking, but rather its "redness". 2 is a more appropriate statement of cause and effect, since it is the redness or non-redness of the stimulus which is, based on the variety of conditions we could submit the pigeon to, the real "difference-maker". There are situations where the stimulus is not scarlet, yet the pigeon does indeed peck – when the stimulus is another shade of red.

Selecting the correct causal claim can be approached by the application of the notion of proportionality: causes should be proportional to their effects, meaning that a statement of a cause should not contain excessive detail, nor omit necessary detail (Woodward is here drawing on Yablo (1992)):

> "(P) There is a pattern of systematic counterfactual dependence (with the dependence understood along interventionists lines) between different possible states of the cause and the different possible states of the effect, where this pattern of dependence at least approximates to the following ideal: the dependence (and the associated characterization of the cause) should be such that (a) it explicitly or implicitly conveys accurate information about the conditions under which alternative states of the effect will be realized *and* (b) it conveys only such information – that is, the cause is not characterized in such a way that alternative states of it fail to be associated with changes in the effect." (Woodward 2010 pg. 298).

So, the possible states of scarletness (scarlet or non-scarlet) are not depended upon by the possible states of pecking (pecking or not pecking). Conversely, possible states of the pigeon's pecking do depend on the possible states of the stimulus' redness. Example 2 provides the required "accurate information about the conditions under which alternative states of the effect will be realised". It also excludes scarletness since this does not provide said accurate information (also fulfilling (b)). Hence redness fulfils the criterion (P), while scarletness does not. This provides a good guide to the formulation of statements of causal relevance – example 2 is much preferred to 1, since it identifies cause and effect better by eliminating those details which violate (P).

Once again, I argue that (M2) comes out on top over (M1) when it comes to satisfying proportionality. The attractor landscape "*displays* or *exhibits* a pattern of dependence" (Woodward 2018, pg. 3, emphasis in original) between causes and effects in the way that a mechanistic story does not. The dynamical model displays and exhibits how the cell's phenotype depends on the stability of its current state, as well as previous and potential future states. The canalisation and perturbation of this dynamical system is therefore proportionate to changes in the cell's phenotype.

Meanwhile (M1) does not relay us accurate information on why certain states of the proposed cause – GATA1 transcription – lead to certain states of the system. Indeed there are a variety of situations where GATA1 transcription will not induce the erythroid fate and situations where (under the same background conditions) other events will induce the erythroid fate. The perturbation may be insufficient, the ridges surrounding the progenitor state too high, and so on.

On both counts – specificity and proportionality – it seems the best candidate cause is described by (M2). Indeed, this is exactly the kind of interpretation that seems prevalent

in the scientific literature and a natural way of speaking about dynamical models of cell fates. For instance, Ferrell (2012) in a review of the role of Waddington's model in cell differentiation emphasises the ultimate role of the attractor landscape in determining fates:

> "Although it was natural to assume that the induction stimulus acts by increasing the value of [a transcription factor] I could have alternatively made the stimulus act through any of the other parameters…Would this alter the conclusion that cell-fate commitment occurs as a result of the disappearance of a valley at a saddle-node bifurcation? The answer is no. No matter how I choose to have the inductive stimulus affect the model, the result is the same." (Ferrel 2012 pg. R461)

Whichever parameter one intervenes on – changing rates of auto-stimulation or mutual inhibition, of transcription levels, etc. – the causal import nevertheless lies with the dynamics of the system, with the attractor landscape. Conspicuous by its absence in these scientific discussions is much concern about how to render the causal story described by these dynamical models down into an underlying mechanism. Rather, in theoretical discussions cell fates are equated with attractors, such that changes to these attractors correspond to changes in cell fate phenomena (Enver et al 2009; Davila-Velderrain et al 2015; Huang 2012; Moris et al 2016). My point here is that the interpretation I offer fits comfortably with the way scientists talk about models of cell fates.

This is not to say that scientific talk, which naturally does not always map directly onto how philosophers of science or biology talk, is decisive to how we should formulate a mode of explanation. On the contrary, I wish to pre-empt the claim sometimes advanced by mechanists, apparently originating with Craver (2007), that the mechanistic mode of explanation most closely reflects how scientists think about nature and experiment on it,

and hence has some especially pragmatic justification. A dynamicist account evidently aligns just as well with the scientific literature at least as far as cell fates are concerned.

5. The 3M Response

The anticipated response from mechanists is that even if attractors meet the interventionist criteria, no such model of cell fates can stand alone as a causal explanation. This is because – according to the 3M requirement – the dynamical model ought to be mapped to an underlying mechanistic model, which is what really provides the causal power and describes the causal structure of cell fates. In other words, 3M is a claim about reduction – causal claims about dynamical models reduce down to causal claims about the underlying mechanistic details. Claims about attractors acting as difference-makers are (if we accept 3M) reducible to claims about difference-makers within the GRN architecture.

However, Kaplan & Craver (2011) stress that 3M is not intended to be rigidly applied, and accordingly set up 3M inclusive of an assumption that the requirement can be defeated, with some caveats:

> "Like all default stances, 3M is defeasible. However, those who would defease it must articulate why mechanistic styles of explanation are inappropriate, what nonmechanistic form of explanation is to replace it, and the standards by which such explanations are to be judged." (Kaplan & Craver 2011, pg. 603)

In taking up this challenge, the first point is likely to be the most contentious. Given the breadth of its application and ambitions, putting a hard limit on the reach of Mechanism is a difficult proposition. However, one uncontroversial starting point is this: Mechanism ends where it is unable to provide explanations. As we have seen, comprehensively

uncovering the causal structure underlying a phenomenon (or at least, being in the process of doing so) is a requirement for explanation for mechanists. As a consequence, if mechanistic models can't get at causal structure, it follows that this would indicate a situation where mechanistic explanation is inappropriate.

Cell fates are an example of the causal structure underlying a phenomenon being accessible to dynamical models. On the other hand, it is inaccessible to mechanistic models. This is because the difference maker is part of the system's dynamics – an attractor – that can only be described in the context of a dynamical model. For example, the progenitor cell fate which makes the difference to the initiation of differentiation "…is "dynamically" defined, namely, as a metastable state in between two neighboring attractors of the prospective differentiated states…" (Huang et al 2007 pg. 699). Hence a description of the mechanism underlying cell fates (the proteins involved, the GRN architecture, etc.) will not get at the difference maker to the cell fate, and consequently there is no clear mechanistic analogue of this attractor, no entity or collection of entities that corresponds coherently to this dynamical feature.

To entertain the notion, persisting with 3M in this case would presumably involve trying to locate the physical components that underlie abstract dynamical features like attractors. This seems like a difficult proposition, since these features don't appear to have a direct relationship to any component of the system – attractors and basins of attraction result from a network of dynamical activity involving many thousands of working parts. To pursue this option necessitates a description of a brute amalgamation of all the physical components associated with these features of the dynamical model – the thousands of genes, the transcription of proteins, etc.

But in this scenario all of the explanatory work would still be done by the dynamical model, since it provides the relevant details about causal relations/difference makers. Identifying physical features associated with a model doesn't entail those features or components are necessarily providing causal detail. It's hard to see how a model like this could be anything other than a bona fide dynamical explanation, since all the explanatory work would be done by the dynamical model.

If anything, the idea of adding in excess mechanistic detail seems to run directly against mechanist's own standards of explanation, as well as the general rationale for these standards. Craver & Kaplan (2018) for instance discuss the notion of completeness as a benchmark of good causal explanation, where a description includes all the causally and constitutively relevant features of the world. Equally critical is the exclusion of irrelevant detail, keeping out those features of the world that are neither causally nor constitutively relevant to the phenomenon.

It is, on the mechanist account, counter to the purposes of a good causal explanation to include features of the world that do not contribute to the unveiling of causal structure. Hence if the addition of further mechanistic detail contributes nothing to understanding the causal structure of cell fates, then it seems that by mechanists' own standards of explanatory completeness 3M is, in this case, counterproductive.

The second requirement for defeating 3M – articulating a non-mechanistic account of explanation – can be dispensed with here more easily. Both Meyer (2018) and van Eck (2018) provide interventionist-based accounts of dynamical explanation – the former outlined earlier in Section 2 – that share many (if not most) of the assumptions about explanation that motivate the mechanist account. This kind of explanation follows the interventionist method of locating difference-makers and developing counterfactual-

supporting invariant generalisations. Further details require significant fleshing out, but these accounts at least offer a foundation to build on.

The third point – what kind of standards would dynamical explanation be judged by – provides the most scope for interpretation. I take this reference to standards to mean, roughly, what is required for a mere description to transition into an explanation. Much like mechanistic explanation, the development of counterfactual-supporting invariant generalisations is the basis of providing causal explanations. Similarly, an appeal to the accounts of dynamical explanation offered by Meyer (2018) and van Eck (2018) ought to be considered at least a good starting point for this project.

One lingering and related issue concerns the capacity for a dynamical model to "fit" onto the real-world system it is supposed to be explaining. Woodward (2017) details this concern, claiming the HKB model of bimanual coordination fails to be explanatory because it underspecifies what the real-world targets of the causal dependency relations revealed by the model actually are. To be clear, Woodward (2017) explicitly denies that this is necessarily a call for an associated mechanism (or a general critique of the possibility of dynamical explanation) but rather for more detailed specification as to what the features of the world purportedly being explained by the model are. Meyer (2018) differs on this point, arguing that it seems fairly clear what the HKB model explains – a cognitive system's exhibition of bimanual coordination.

To expand on this point, I think it is worth looking at why mechanistic explanations do not typically face this problem of fit. For one, mechanistic models seem intuitively easier to fit onto the world – after all, mechanistic models are supposed to (to some extent, with necessary abstractions and idealisations) resemble the target mechanism-in-the-world (Craver 2007). But more formally, a mechanistic model fits onto features of the world and

explains them when the counterfactual dependencies it describes resemble counterfactual dependencies actually obtaining in the world. Dynamical explanations could do the same: in the cell fates example, we have a model that describes counterfactual dependencies which the investigators show also obtain in the GRNs of real FDCP-mix cells (Huang et al 2007). Hence the dependencies in the model resemble those in the real-world system – and fit is achieved.

Another important aside here concerns the relationship between mechanistic and dynamical explanations. So far this paper has not grappled with problems of integration or compatibility explicitly. I would argue that the specific characterisation of the phenomenon – the explanatory question being asked – will play a large role in how these modes of explanation hang together. A question specifically about the difference-maker to cell fates appears to require what van Eck (2018) calls a "pure dynamical model", one that is non-mechanistic. We are really only interested in the dynamics causing one outcome to obtain over another, and include mechanistic details (the GRN) only as background conditions. All the relevant causal detail is contained in the dynamical model.

I leave open the question of a combination of the two, which seems a substantial discussion requiring its own treatment. I have only discussed one example from one subgenre of science, and hence it is plausible that some questions require a different model or combination of models to fully encompass the phenomenon.

6. Conclusion

In this paper I have argued that non-mechanistic, causal dynamical explanations are viable. I have provided reasons, rooted in Woodward's interventionist account, for thinking that dynamical models do in fact describe the causal structure of the world independently of mechanisms, using the example of cell fates. I have argued that

dynamical models can describe more specific and proportional causes than mechanistic models of the same phenomena, and hence are the better descriptors of causal structure in these cases. I have also defended against the 3M criterion, showing how in the case of cell fates, adhering to 3M is counter-productive to causal explanation and need not be observed.

References

Baumgartner, M., & Gebharter, A. (2016). Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731–756. https://doi.org/10.1093/bjps/axv003

Chemero, A., & Silberstein, M. (2008). After the Philosophy of Mind: Replacing Scholasticism with Science*. *Philosophy of Science*, *75*(1), 1–27. https://doi.org/10.1086/587820

Chirimuuta, M. (2017). Explanation in Computational Neuroscience: Causal and Non-causal. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axw034

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376. https://doi.org/10.1007/s11229-006-9097-x

Craver, C.F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

Craver, C.F. & Kaplan, D.M. (2011). Towards a mechanistic philosophy of neuroscience. In S. French & J. Saatsi (Eds.), *Continuum Companion to the Philosophy of Science* (268-292). London: Continuum.

Craver, C. F. & Kaplan, D. M. (2018). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axy015

Davila-Velderrain, J., Martinez-Garcia, J. C., & Alvarez-Buylla, E. R. (2015). Modeling the epigenetic attractors landscape: toward a post-genomic mechanistic understanding of development. *Frontiers in Genetics*, *6*. https://doi.org/10.3389/fgene.2015.00160

Enver, T., Pera, M., Peterson, C., & Andrews, P. W. (2009). Stem Cell States, Fates, and the Rules of Attraction. *Cell Stem Cell*, *4*(5), 387–397. https://doi.org/10.1016/j.stem.2009.04.011

Ferrell, J. E. (2012). Bistability, Bifurcations, and Waddington's Epigenetic Landscape. *Current Biology*, *22*(11), R458–R466. https://doi.org/10.1016/j.cub.2012.03.045

Glennan, S. & Illari, P. (2018a). Introduction. In S. Glennan & P. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. Abington: Routledge.

Glennan, S. & Illari, P. (2018b). Varieties of mechanisms. In S. Glennan & P. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. Abington: Routledge.

Graf, T., & Enver, T. (2009). Forcing cells to change lineages. *Nature*, *462*(7273), 587–594. https://doi.org/10.1038/nature08533

Huang, S. (2009). Reprogramming cell fates: reconciling rarity with robustness. *BioEssays*, *31*(5), 546–560. https://doi.org/10.1002/bies.200800189

Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? *BioEssays*, *34*(2), 149–157. https://doi.org/10.1002/bies.201100031

Huang, S., Eichler, G., Bar-Yam, Y., & Ingber, D. E. (2005). Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. *Physical Review Letters*, *94*(12), 128701. https://doi.org/10.1103/PhysRevLett.94.128701

Huang, S., Guo, Y.-P., May, G., & Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, *305*(2), 695–713. https://doi.org/10.1016/j.ydbio.2007.02.036

Huang, S., & Ingber, D. E. (2007). A Non-Genetic Basis for Cancer Progression and Metastasis: Self-Organizing Attractors in Cell Regulatory Networks. *Breast Disease*, *26*(1), 27–54. https://doi.org/10.3233/BD-2007-26104

Jaeger, J., & Monk, N. (2014). Bioattractors: dynamical systems theory and the evolution of regulatory processes. *The Journal of Physiology*, *592*(11), 2267–2281. https://doi.org/10.1113/jphysiol.2014.272385

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339–373. https://doi.org/10.1007/s11229-011-9970-0

Kaplan, D. M. (2015). Moving parts: the natural alliance between dynamical and mechanistic modeling approaches. *Biology & Philosophy*, *30*(6), 757–786. https://doi.org/10.1007/s10539-015-9499-6

Kaplan, D. M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective*. *Philosophy of Science*, *78*(4), 601–627. https://doi.org/10.1086/661755

Krickel, B. (2017). Making Sense of Interlevel Causation in Mechanisms from a Metaphysical Perspective. *Journal for General Philosophy of Science*, *48*(3), 453–468. https://doi.org/10.1007/s10838-017-9373-0

Meyer, R. (2018). The Non-mechanistic Option: Defending Dynamical Explanations. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axy034

Moris, N., Pina, C., & Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*, *17*(11), 693–703. https://doi.org/10.1038/nrg.2016.98

Ross, L. N. (2015). Dynamical Models and Explanation in Neuroscience. *Philosophy of Science*, *82*(1), 32–54. https://doi.org/10.1086/679038

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Scholz, J. and Kelso, J. (1989). A Quantitative Approach to Understanding the Formation and Change of Coordinated Movement Patterns. *Journal of Motor Behavior*, *21*(2), 122–44.

Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the Rest of Cognitive Science. *Topics in Cognitive Science*, *3*(2), 425–437. https://doi.org/10.1111/j.1756-8765.2011.01143.x

van Eck, D. (2018). Rethinking the explanatory power of dynamical models in cognitive science. *Philosophical Psychology*, *31*(8), 1131–1161. https://doi.org/10.1080/09515089.2018.1480755

Waddington, C. H. (1957). *The Strategy of the Genes*. London: George Allen & Unwin.

Waters, C. (2007). Causes that make a difference. *The Journal of Philosophy*, 104(11), 551-579. https://doi.org/10.5840/jphil2007104111

Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2008). Mental causation and neural mechanisms. In J. Hohwy and J. Kallestrup (Eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation* (218–262). Oxford: Oxford University Press.

Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318. https://doi.org/10.1007/s10539-010-9200-z

Woodward, J. (2015). Interventionism and Causal Exclusion. *Philosophy and Phenomenological Research*, *91*(2), 303–347. https://doi.org/10.1111/phpr.12095

Woodward, J. (2017). Explanation in Neurobiology: An Interventionist Perspective. In Kaplan, D.M. (Ed.), *Explanation and Integration in Mind and Brain Science*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199685509.001.0001

Woodward, J. (2018). Explanatory autonomy: the role of proportionality, stability, and conditional irrelevance. *Synthese*. https://doi.org/10.1007/s11229-018-01998-6

Yablo, S. (1992). Mental Causation. *The Philosophical Review*, *101*(2), 245-280. https://doi.org/10.2307/2185535