Novel Explanation in the Special Sciences: Lessons from physics

Eleanor Knox

Kings College London

30th March 2017

Abstract

This paper aims to understand how recent discussion of novel and robust behaviour in physics might be applied in biology and other special sciences. In particular, it looks at the prospects for extending an account of novel explanation to biological examples. Despite the differences in the disciplines, the prospects look good, at least when we look at a biological example in which a certain kind of reduction is possible.

I. Introduction

The study of intertheoretic relations has been undergoing something of a renaissance in both philosophy of physics and philosophy of biology. Both disciplines are moving away from the battle lines set in the mid twentieth century; the field is no longer defined by the debate over reductionism. The focus has moved instead to features of intertheoretic relations that could be present even when the formal relationship between levels is well understood. In the philosophy of physics, recent discussions of novel and robust physical behaviour[1] have shed light on the sense in which physical phenomena can class as autonomous, while in biology, a discussion of biological robustness (and, more generally, causal mechanisms) is helping us to understand how biological systems maintain their behaviour through underlying change.

This paper is an attempt by one all-too-biologically-ill-informed philosopher of physics to make some connections between these two areas of development. In particular, I'm interested in whether an account of novelty as novel explanation offered in two recent papers (Knox 2016; Franklin and Knox 2017) might be applicable to biological cases. What follows will look at one biological example, and argue that the prospects for extending the account look good, despite differences in the nature of the two fields.

To set the scene, it is worth saying something about how the connection between recent literatures in the two disciplines does *not* work. One way of characterising the discussion of novel and robust behaviour in physics is as discussing *emergence.* Biological robustness may be thought of as a type of multiple realisation.[2] One might then think that the connection between discussions in physics and biology was obvious: if emergence entails the failure of reduction, as does multiple realisation, then the two literatures might very well be getting at the same thing. But I accept neither of the statements on which this link is premised. Emergence in the philosophy of physics is used more in the physicist's sense than in the philosopher's and is *not* generally taken to imply the failure of reduction. And there are good

[1] Discussion of the importance of novel and robust behaviour originates with Jeremy Butterfield (2011a; 2011b).

[2] Although we'll see in section IV that there is reason to at least qualify this claim.

reasons enumerated elsewhere[3] not to accept the Putnam-Fodor argument from multiple realisation to the failure of reduction.

Instead, the connection between recent work in biology and novelty in physics (or at least, my account of it) is more subtle. Biological robustness[4] is often a feature of properties that lead to interesting explanatory abstractions. And, in many cases, the explanatory value of these abstractions is not obvious when the description is translated into the language of some lower level science. This explanatory novelty of a higher level science gives us one sense in which higher level science might be autonomous of more detailed description even where reductions are available.[5]

Section II will discuss the idea above as it applies in physics, illustrating the case with a simple example – that of normal modes of a simple harmonic oscillator. In this case, a change of variables allows new explanatory abstractions. But it's not immediately obvious that the relevant features carry over to biological cases; for one thing, the example given involves laws formulated entirely in mathematical language.

Section III will discuss a biological example – the characteristic output patterns of the stomatogastric ganglion of a lobster (which is a simple neuronal network). Interestingly, this example has much in common with the normal modes case, despite being a paradigmatically biological system, and one not entirely characterised by mathematical laws. Characteristic patterns of the stomatogastric ganglion play a role in explanatory abstractions that may seem mysterious at the level of the neurons themselves.

But one might think that a system like the stomatogastric ganglion exhibits a deeper novelty than non-biological systems. This is because biological traits characteristically are said to possess a particularly strong form of robustness; biological systems evolve to exhibit certain traits even under diverse conditions. Section IV will discuss this robustness alongside related ideas in physics, and conclude that it is the same kind of robustness that we find in physics. As a result, our two examples are analogous in important respects, and many of the morals drawn in physics may well go over to a subject like biology.

II. Explanatory novelty in physics

In two recent papers (Knox 2016; Franklin and Knox 2017), Alex Franklin and I offer an account of *novelty* in physics that claims that one level of description in physics can be novel relative to another just as a result of the explanatory abstractions that are made available when we change the variables that we use to describe a system. This account of novelty (which I'll explain further shortly) is intended to fill a lacuna in an interesting account of emergence in physics offered by Jeremy Butterfield (2011a; 2011b). Butterfield analyses emergent behaviour of a system as behaviour that is "novel and robust relative to some comparison class"(Butterfield 2011b p.1065), for example, relative to characteristic behaviour at smaller

[3] As will become clear, I find the arguments given in (Wilson 1985) particularly persuasive.
[4] For a discussion of biological robustness see (Boone 2016) . What follows here relies heavily for its biological ideas on this piece of work, which first introduced me to the stomatogastric ganglion.
[5] Ours is not the only current work focussing on the importance of explanatory abstractions at different levels. For an interesting, related, account see (Haug 2011).

scales. This account promises to offer an account of the kind of emergence that physicists discuss – an emergence that is often explicitly held to be compatible with reduction.[6]

This account requires an explanation of the relevant senses of robustness and novelty. Butterfield takes robustness to mean the maintenance of some higher level behaviour despite perturbations in the lower level description. This phenomenon, which will be discussed in detail in section IV, has been relatively well-explained in recent philosophy of physics literature. But providing an account of novelty is more difficult, particularly if the kind of novelty we seek must be explicitly compatible with reduction. Butterfield and others[7] look to asymptotic limiting relations between theories to reveal a relevant kind of novelty; in (Franklin and Knox 2017), Alex Franklin and I argue that this kind of novelty will not extend to important cases.

Instead, we think many cases are characterised by *explanatory novelty* that arises when we change the variables or quantities with which we describe the system and then perform explanatory abstractions based on the new variables. This idea is premised on an account of explanation on which abstraction leads to *better* explanation; it assumes that Peter Railton (1981) is wrong when he claims that a perfectly detailed explatory text is *ideal*. Instead, the account at hand builds on numerous contemporary accounts of explanation[8] that focus on the importance of eliminating irrelevant detail to our explanatory practice. If we accept that abstracting away from detail is a central part of offering a good explanation, then a change of variables can make available better explanations, because it offers the opportunity for new abstractions. Where this happens, the new description has *novel explanatory value.*

To see how this kind of novelty might arise, consider a simple physics example (also discussed in Franklin and Knox 2017):

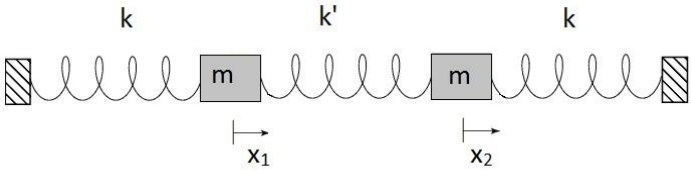

Fig. 1: Two masses on springs.

Figure 1 shows two particles of equal mass m oscillating on springs with constants k and k'. This system obeys the following equations of motion:

[6] I will say more about reduction in section III, but for the time being, I mean something like Nagelian reduction (deduction of the higher level theory from a lower level theory) with a relatively liberal notion of what can constitute a bridge law.

[7] For example, Bob Batterman (2002a; 2009; 1995; 2010; 2002b; 2005). But much of Batterman's work explicitly endorses anti-reductionism, so the sense in which he thinks higher level theories are novel is clear.

[8] For example, those given in (Strevens 2008) and (Batterman 2002b). I intend what follows here to be independent of the actual model of explanation offered, as long as whatever model we choose agrees on the importance of abstraction. While Strevens offers an account of the importance of abstraction to *causal* explanation, Batterman moves between causal explanations and other kinds.

$$m\ddot{x}_1 = -kx_1 - k'(x_1 - x_2) \tag{1}$$

$$m\ddot{x}_2 = -kx_2 - k'(x_2 - x_1) \tag{2}$$

In order to solve these equations, we change our variables to 'normal mode variables':

$$\eta_1 = x_1 + x_2, \tag{3}$$

$$\eta_2 = x_1 - x_2, \tag{4}$$

With this change of variables, the equations of motion decouple and become:

$$m\ddot{\eta}_1 = -k\eta_1 \tag{5}$$

$$m\ddot{\eta}_2 = -(k + 2k')\eta_2. \tag{6}$$

These are simple harmonic oscillator equations, and have familiar solutions:

$$\eta_1 = 2A_s \cos\left(\sqrt{\frac{k}{m}}t + \phi_s\right) \tag{7}$$

$$\eta_2 = 2A_f \cos\left(\sqrt{\frac{k + 2k'}{m}}t + \phi_f\right) \tag{8}$$

With this change of variables we've characterised the system in terms of its two modes of oscillation, instead of in terms of the displacements of the two masses; one mode corresponds to the two masses oscillating in tandem, and the other to oscillation with the masses moving in opposite directions. This change makes calculations much easier, but it may also enable us to give a better explanation for certain phenomena.

Suppose we place a light in the central spring that lights whenever the spring is compressed beyond a certain point, and set the system going in its second normal mode η_2. If we would like to explain the frequency of the light flashing, then the first normal mode η_1 and its associated equation are irrelevant to our explanation – we can explain the phenomenon in terms of one equation and a single variable. But any explanation in terms of the original variables will involve two variables;[9] our simpler and better, abstracted explanation is only available after the variable change. It's precisely because the abstraction at the higher level 'cuts across' the variables at the lower level that it provides explanatory value not available at the lower level.

For those who believe that abstraction leads to better explanations, this should be enough to establish that variable changes can come with a change in explanatory value. But we might think that the *novelty* here is relatively weak; we can translate the explanation back to the displacement variables, even if the reasons for the choice of a particular presentation of the displacement variables isn't obvious without the normal modes variable change. But if we

[9] In fact, without changing the variable, any explanation will also likely involve two equations; it's only the variable change that makes it obvious *how* we decouple the equations. Of course, once we have the decoupled equations in the new variables, we can always translate back into displacement variables.

move to a more complex example,[10] with many more than two displacement variables, then back translations will become less and less illuminating, and, in some cases, where information is lost in the variable change, the back translation will not be possible at all. Novelty, perhaps, admits of degrees – our normal modes variable change above allows for explanations that display a weak kind of novelty, but in more complex examples, a stronger form of novelty may emerge.

If this account is correct, then moves between different levels of description in physics will often lead to novel explanations because they will almost always allow for new abstractions. But might this kind of novelty also arise in the special sciences? The account so far has been highly mathematical, and the changes in descriptive quantities described are changes of variable in the precise mathematical sense. It's not obvious that this account could go over to, say, biology, where characteristic quantities and properties are not always connected by mathematical relationships.

That said, explanatory abstractions are just as important in the special sciences as in physics; choosing the right quantities and properties with which to describe a system (and omitting irrelevant information) is very much the name of the game when describing complex systems. So one might expect *something* like the above to apply outside physics.

III. An example from biology

This section will discuss a biological (and neurophysiological) system which has several features that are similar to the normal modes example above. The stomatogastric ganglion of a lobster is a simple neuronal network that controls the digestive function of the animal. It's of particular interest because, although the neurons themselves are much like neurons in any animal (including mammals), the network is simple enough to be straightforwardly modelled and understood. This means that it's not only an interesting example for neuroscientists, but also for philosophers interested in understanding how descriptions at the level of individual neurons relate to descriptions at the level of network-wide outputs.

The stomatogastric ganglion is a system of 30 neurons located on the wall of the digestive tract of the lobster. It contains two central pattern generators (CPGs): that is, generators of characteristic signals that control basic digestive behaviours. The pyloric CPG controls peristaltic motions that pass food down the gut. Of interest to us here, however, is the gastric CPG, which controls the motion of three internal teeth.[11] The gastric CPG typically generates two different patterns, which move the teeth in different ways: Type 1 patterns cause the three teeth to squeeze together simultaneously and Type 2 patterns cause the two lateral teeth to move in opposition to the medial tooth in a cut and grind motion.

Figure 2 shows recordings of muscle contractions in live lobsters. The gastric mill CPG is comprised of 10 motor neurons that stimulate muscles, and just one connecting neuron, so muscle contraction is an accurate proxy for groups of neurons firing.[12] The top two signals show the muscles that contract and retract the two lateral teeth, and the bottom one the muscle that retracts the medial tooth. Squeeze chewing happens when the contraction is simultaneous, as in B_1, and cut and grind when the muscle contraction is out of phase, as in B_2.

[10] One could, for example, consider the normal modes of vibration in the atoms of a macroscopic crystal, with displacement variables on the order of 10^{26}. Franklin and Knox (2017) discusses just this example.
[11] Interestingly, lobsters have teeth located well past what we'd think of as the mouth (the mandibles).
[12] For more on this, and a general overview of the function of the gastric mill, see (Heinzel 1988).

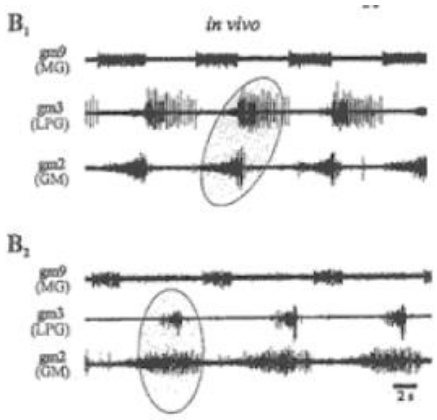

Fig. 2: Gastric Mill output patterns as measured in live lobsters. Pattern B_1 leads to the squeeze mode of chewing, while pattern B_2 leads to the cut and grind mode. Reproduced with permission from Springer Nature from (Combes, Meyrand, and Simmers 2002) p.582.

Just as in our normal modes example above, the two modes of tooth operation are characterised by two individual patterns being in or out of phase with one another. And just as in the physics example, there are higher level phenomena that are better explained by appealing to the mode of operation than to a component level description: if we want to explain how lobsters are capable of digesting a particularly tough meal, for example, we might want to explain how the meal stimulates the anterior gastric receptor that causes type II CPG output.[13] There is an obvious similarity, at least superficially, to explaining the flashing light in terms of just one normal mode of oscillation in section II.

Ought this similarity of our two examples to lead us to conclude that the gastric mill CPG demonstrates the same kind of explanatory novelty exhibited by our normal modes example? Ultimately, yes, but one might have doubts; there are important differences too. The most important is the obvious one: the kinds of formal change of variables involved in the physics example would never be made explicit here. Indeed, despite the relatively quantitative description of CPG output patterns via diagrams like those in Figure 2, one wouldn't tend to characterise the two output patterns via variables at all. A typical discussion would appeal to general features of the patterns rather than a new class of variable and its governing equations.

Why might this difference, the lack of a complete quantitative analysis at the CPG level, be important here? In the normal modes example, the availability of a reduction was not in question: we can think of equations 3 and 4 as expressing bridge laws in the full Nagelian sense. What follows thereafter is simple deduction.[14] The particular kind of explanatory novelty in that example was one that was not only compatible with reduction but *required* a

[13] In the interests of full disclosure, I should note that I haven't been able to find information about exactly what kind of meal triggers type II chewing; it's an oddity of this particular field (one which tells us something about the particular reductive aims of the field) that studies of lobster chewing are much more tightly focussed on modelling neuronal systems than they are on lobster behaviour itself.

[14] The only objection to seeing the normal modes case as a case of reduction would be that the two descriptions are too closely connected to count as distinct theoretical levels at all. But the point here is that novel explanations can be provided even by such closely related descriptions.

mathematical relationship between higher and lower level variables. Without a mathematical characterisation of the CPG, perhaps we should doubt that the gastric CPG is really reducible to a description in terms of individual neurons.

Needless to say, whether we classify the gastric CPG description above as reducible will depend on our definition of reduction. If we mean full Nagelian reduction – the kind of deduction with the aid of bridge laws that we had in the normal modes case – then we may question whether a reduction exists here, at least in practice. One might think that patterns B_1 and B_2 in Fig. 2 are not always characterised in sufficiently explicit mathematical terms to make a deduction available (even before we worry about whether the bridging principles used in the deduction meet whatever requirements we have of bridge *laws*).

But actual practice in this field belies this view of the CPG's reducibility. Why do neurophysiologists who would otherwise not be interested in lobsters (except on a restaurant menu) study the stomatogastric ganglion? They do this precisely because the relationship between theoretical levels here is close: we have a very good understanding of how the action of individual neurons contributes to the production of the overall pattern; the study of the stomatogastric ganglion often involves measuring the activity of individual neurons and modelling the whole system as a circuit, and this modelling is very successful indeed. We then characterise the relevant CPG patterns in terms of neuronal activity.

Now, of course, we don't usually write down a set of CPG variables that can be described as a function of underlying neuronal variables. Instead, studies describe CPG outputs in largely qualitative terms; the patterns of Fig. 2 are most easily described by relative burst orders, rather than more formally. But it would be strange to think that finding appropriate variables is impossible in principle;[15] synchronicity and order of a pattern are easily mathematized if we allow ourselves some approximation. One can classify the outputs of each neuronal burst via variables that take simple binary on-off values, and then describe the relevant patterns via relative sequences of these values. It's certainly a simple matter to deduce the CPG pattern from basic information about the neuronal bursts. It's also a simple matter to characterise the neuronal bursts in terms of aggregate actions of the group of neurons; there is no mystery here as to how the CPG output relates to activity at the level of neurons. Likewise, the relationship between neural and output is understood. When studies such as Prinz et al. (2004) use computer modelling to predict CPG output patterns based on precise input values, it's exactly this feature that they're exploiting. The study in question ran 20,250,000 simulations of neural network behaviour and classified these as corresponding to different CPG modes. Although the method of classification isn't listed, it certainly wasn't manual! Programming a

[15] In a 1985 paper that has received less attention than it deserves, Mark Wilson criticises anti-reductive arguments, in part, because they underestimate our ability to find appropriately defined physical variables for any property of interest. Physics, Wilson points out, has been far more inventive in coming up with appropriate variables than bad philosophical examples (like temperature as mean kinetic energy) would have us believe. With access to a full mathematical toolbox, it's possible to find a variable that helps to carve the space of states in any way that we desire.

computer to simulate the network and recognise the relevant patterns in the data requires characterising the neural output in the kind of quantitative way outlined above.

Should we then say that a reduction is possible? If one means Nagelian reduction, this will depend on the restrictions that we put on our bridge laws. However we cash out the exact characterization of the CPG pattern, these patterns will be realised by many values of the neuronal inputs. Some conceptions of reduction will exclude deductions using these bridging principles because the principles involve problematic multiple realisability.

I don't plan to rehash the literature on multiple realisation here. For what it's worth, the above view strikes me as close to a *reductio* of a certain view of reduction; we can easily model all aspects of the CPG pattern, and we have a clear understanding of every step in the move from neuronal description to output pattern. If this doesn't count as a reduction, then reduction in the philosopher's sense has shifted too far from its usage in the scientific community. Much better instead to accept that multiple realisability, at least of this kind, is compatible with reduction.

But, as it happens, my arguments do not turn on the details of Nagelian reduction; although I'll refer to reduction in what follows, advocates of a strong Nagelian account may substitute a weaker term of their choice. Reduction was of interest to us here because obvious failures of reduction would lead to obvious cases of novelty; where one descriptive level is unrelated to another, it can be counted as novel without the need for the kind of account of explanatory novelty given here. But the relation between neuron firings and CPG outputs here leave no room for this strong kind of novelty; whether or not we call the relationship between levels a reductive one, it's sufficiently well understood that only a weaker notion of novelty can apply here.

So let us suppose then that, in principle, the CPG patterns can be characterised in terms of variables that are some complex, approximated function of the lower level variables. In this case, what's going on here in terms of explanation looks much like the normal modes case. When we appeal to just the type 2 pattern, we are abstracting away from the variable associated with the type 1 pattern, and by doing so we produce a better explanation. Thus, if the normal modes example involves novel explanation, this one does as well.

All of this is not to say that the significant difference between a biological example and a very simple physical one are not important here – far from it. For one thing, even with a very simple network of 11 neurons, this example is hugely more complex than the masses on springs. For another, any relation between variables will involve approximation and idealization to a greater extent than the masses on springs. But both of these features make the novelty in question stronger, because they make the value of the *choice* of explanatory information at the CPG level even less obvious from the neuronal level.[16]

Thus far, I have said little about robustness – it will be the topic of the next section. But, before we move on to the topic, I'd like to clarify a difference between two ways in which descriptions

[16] The idea here is that the choice of explanatory information can be non-obvious from the fundamental level even when the relationship between levels is well understood. This is the lesson of the normal modes example: when we focus on the description at the level of the two displacement variables, the interest and importance of the normal modes variables is not apparent.

at different levels might be related (two different characteristics that inter-level bridge laws can have, if you will). These will turn out not just to be relevant to the discussion here, but also to the issue of robustness.

First, we might change our description simply by changing the level of detail that we're interested in. This kind of relationship between quantities involves abstracting away from detail during the quantity change itself (not just afterwards via cross-cutting abstractions). We would not ordinarily call the description that results from this process alone novel – the kind of 'zooming out' that happens here seems to be a paradigm of classic, simple reduction.

Second, quantities can be related via a change of variables as expressed in equations 3 and 4. In this kind of case, the system could be characterised by the same number of variables at one level that it is at the other, and there might be no loss of detail, or abstraction, when we move from one class of variables to the other. But this kind of variable change does allow for novel explanatory value of one level relative to another if we then perform explanatory abstractions based on the new variables – it is just this kind of variable change that leads to 'cross-cutting' abstractions.

Unlike our normal modes example, most changes in descriptive quantity involve both kinds of relationship. For example, when we move from a description in terms of the individual molecules of an ideal gas to one in terms of temperature, we must both zoom out (temperature is only theoretically useful on relatively long length scales), and take a function of the underlying variables (in this case, the mean). Our gastric mill CPG example is no exception; a large number of neuronal variables feed into just two modes of operation at our (simplified) CPG level, so the number of variables needed to characterise the system will drop dramatically. At the same time, mere coarse-graining won't work if we're interested in characterising the CPG output patterns.

IV. Robustness

The last two sections aimed to establish that, even if the CPG pattern is reducible to the neuronal description, there is still a sense in which it is novel: it allows for novel explanatory abstractions and thus leads to novel explanatory value. But might this vastly understate the novelty involved in a case like this? Here, robustness enters the scene. One variable is robust relative to some class of lower level variables when that variable remains invariant, or unchanged, over a range of changes to lower level variables. There has been much recent discussion of 'biological robustness'. If this means something much stronger than robustness in the physics literature, we might suppose our biological example to involve considerably *more* novelty than our simple physics example.

I'll argue here that, for the purposes at hand, biological robustness is much like robustness in physics. Let's start with the philosophy of physics literature. As noted earlier, Jeremy Butterfield proposes a definition of emergence as *novel and robust* behaviour, and holds that such behaviour can exist even where reduction is possible. Leaving aside the issue of whether emergence is the right term for this (it's certainly weaker than other definitions), let's agree that the existence of novel and robust behaviour would be of independent interest. What is meant by robustness here? I'll take a higher level phenomenon to be robust just in case it survives perturbations of lower level variables. But we should be careful here: there are two kinds of perturbation we might consider.

First, we might consider perturbations of (small changes to) lower level variables *while holding other lower level variables fixed.* That is, all other things being equal, the value of some higher level variable and the phenomenon it underpins only cares about the approximate value of a lower level variable, and not its exact value. Robustness under this kind of perturbation will be seen when our descriptions at different levels are connected by a change in the level of detail captured. If we get to a higher level description precisely by washing out details at the lower level, the higher level variables will generically be invariant under perturbations of this lower level detail.

But we might also consider perturbations of lower level variables *while permitting changes in other lower level variables.* In this case, the change to one variable is compensated for by the change in other variables. If we look at the normal modes variable η_2 in equation 4, it's obviously possible to compensate for any change to x_1 with a change to x_2. This possibility is a generic feature of functions of more than one variable, although whether the higher level variable's invariance under this kind of perturbation is relevant and useful will depend on the details of the case.

This is not all there is to be said about invariance in physics; it may well be, for example, that taking the limit of some function as a parameter tends to infinity (as we do when we take the thermodynamic limit), yields a particularly strong form of robustness. But this is enough to be getting on with for the biological discussion at hand.

So let us turn to biology. *Biological robustness* is of considerable interest to both biologists and philosophers. In a 2004 paper in Nature, Hiroaki Kitano defines robustness as "a property that allows a system to maintain its functions despite internal and external perturbations". Aside from the characteristically biological mention of 'function' (which I take it tells us something about the *kinds* of variable that biology is interested in), this sounds much like the physics definition. But, quite plausibly, biological systems are said to evolve to exhibit particularly strong forms of robustness – their function can often be maintained over a remarkable range of perturbations.

In the philosophy of biology literature, this talk of robustness has fed into debates on causal mechanisms in interesting ways. An interesting paper by Trey Boone (2016) claims that biological robustness is multiple realisation by another name, and that the stomatogastric ganglion of the lobster exhibits this multiple realisation.

I find little to argue with in Boone's paper itself. Its aim is to build on work by Larry Shapiro (2000) and reframe multiple realisation as a phenomenon relevant to causal explanation, rather than as one directly relevant to Nagelian reduction. For Boone, the fact that some phenomenon is multiply realised tells us something about causal structure at different levels; multiple realisation describes those cases in which a particular function can be realised by more than one causal mechanism. Much of this is very congenial to what I have to say here; my account depends on the idea that explanatory dependencies at different levels need not align, and the causal mechanisms literature is entirely compatible with this idea. But Boone also builds on literature in biology that sees biological robustness as uniquely strong, and presents a link between *biological* robustness and multiple realisability. This might suggest that our gastric mill CPG example differs in important respects from the normal modes example above.

To get a handle on exactly what is meant by biological robustness, let's look again at the stomatogastric ganglion. In a 2004 paper discussed by Boone, Prinz, Marder and Bucher examine the robustness of the stomatogastric ganglion's pyloric CPG output under perturbations of underlying neuronal variables like synaptic strength. Via computer modelling, they conclude that the output of the CPG is maintained under a vast array of very diverse variable values. Moreover, they note that for more or less any given value of an underlying variable, there is a way to tune the other variables such that the pattern is maintained. Assuming that their modelling reflects the behaviour of real neuronal systems, this suggests that systems like the stomatogastric ganglion exhibit a strong breed of robustness.

Some caution is needed here, because the 2004 paper concerns only the pyloric CPG (and, indeed, a simplified model of the pyloric CPG), and I am not aware of similar work for the gastric mill CPG. But let us assume that the result carries over to the gastric mill CPG in real (as opposed to computer-modelled) lobsters. Do we have here a relevant disanalogy with the physics case? Does this system demonstrate a form of biological robustness that prevents us thinking about the relationship between higher and lower level variables in much the same way that we do in physics?

No. The fact that output can be maintained over a large range of lower-level perturbations is exactly what we'd expect if higher level variables exist that are expressible as a function of multiple lower level variables. The perturbations involved are our second kind of perturbation above – ones where we allow other variables to change in dynamically permitted ways. There is nothing about this particular feature of biological robustness that cannot be captured by our account for physical variables. Of course, in the absence of a more precise account of the relationship between biological variables, this argument is suggestive rather than conclusive, but the point here is that the feature of biological systems displayed here – invariance of higher level qualities under perturbations of the second kind – does not force a new kind of novelty on us.

That is not to say there is *nothing* special about biological systems with respect to robustness. All of this is entirely compatible with the idea that biological systems evolve to display strong robustness; it is quite likely that they evolve so that important higher level variables are given by functions of lower level variables that are invariant under relevant changes. It's just that this robustness isn't different in kind to that displayed by physical systems.

V. Conclusions

The upshot of all of this is that the kind of analysis of novel and robust behaviour that I have offered elsewhere for physics goes over rather well to this particular biological example. Plausibly, the relationship between CPG and neuronal variables, if formalised, would show us that CPG level explanations involve the kind of cross-cutting abstraction that we see in physics examples. Aside from the lack of an explicit formulation of this relationship, the distinctively biological features of the system don't impact the analogy.

If the analogy does go through, there is at least one biological system whose higher level description is novel and robust relative to a cell-by-cell description despite being reducible to it. And in this case the novelty is explanatory novelty; explanations given at the higher level will rely on abstractions that look unnatural from the lower level. Another way of saying this is that within the lower level description, relevance relations will not track natural distinctions between variables.

What this offers us is an understanding of novelty even where varia bles are related in such a way that stronger novelty or autonomy might seem impossible. It does not tell us why some variables definable in the lower level are particularly interesting, nor why some explanatory questions are particularly important. A full understanding of scientific levels requires more work on both these questions.[17] Nonetheless, it is helpful to see that novelty and reduction (or something like it) need not be in tension in this biological context.

What should this lead us to say about the special sciences more generally? My example is obviously cherry-picked; it's no coincidence that I chose a system that exhibits two characteristic patterns of output much like the normal modes case. But that similarity was merely to make the connection easier to see. What's really important about this example is the obvious connection between the CPG pattern to underlying variables; it means that the idea of a formalisable relation between higher and lower level variables is plausible.

How far might we wish to extend the conclusions? CPGs govern rhythmic movements in general, including quite complex ones like a horse's gaits. For these more complex systems, the only currently available explanations take place at the CPG level, rather than the neuronal level. But it's hard to see why mere complexity should affect the in principle, rather than in practice, availability of a reductive relation between variables. And perhaps what goes for CPGs in simple neural systems like the spine also goes for brains – who knows?

I am optimistic about the prospects for reduction, at least if we acknowledge that it's a messy, approximative and piecemeal affair. But my aim here is not to argue for reductionism. It is, rather, to argue that higher level descriptions may possess explanatory novelty even where reduction is possible and available. And that lesson, at least, does seem to transfer from physics to the special sciences.

Acknowledgements

I am very grateful to Alex Franklin for numerous discussions that have shaped my thinking on this topic, to Jeremy Butterfield for helpful suggestions on related work, and to Trey Boone for providing me with a good biological example at the 2016 PSA.

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

———. 2002. "Asymptotics and the Role of Minimal Models." *Theq1` British Journal for the Philosophy of Science* 53 (1). 21.

[17] I suspect the answer to both requires an understanding of the way in which some classes of variables play a role in a wide-ranging network of dependencies (causal or otherwise) while others do not. Any answer along these lines will connect to the issues of robustness discussed here.

———. 2005. “Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics.” *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 36 (2). 225–44.

———. 2009. “Idealization and Modeling.” *Synthese* 169 (3). 427–46.

———. 2010. “Emergence, Singularities, and Symmetry Breaking.” *Foundations of Physics* 41 (6).

Boone, Trey. 2016. “Multiple Realization and Robustness” *Unpublished*.

Butterfield, J. 2011a. “Emergence, Reduction and Supervenience: A Varied Landscape.” *Foundations of Physics* 41 (6) 920–59.

———. 2011b. “Less Is Different: Emergence and Reduction Reconciled.” *Foundations of Physics* 41 (6). 1065–1135.

Combes, Denis, Pierre Meyrand, and John Simmers. 2002. “Motor Pattern Switching by an Identified Sensory Neuron in the Lobster Stomatogastric System”, *The Crustacean Nervous* System, K. Weise ed. Springer. 582–90.

Franklin, Alexander and Knox, Eleanor. 2017. “Emergence Without Limits : A Case Study.” *Unpublished*.

Haug, Matthew C. 2011. “Abstraction and Explanatory Relevance; Or, Why Do the Special Sciences Exist?” *Philosophy of Science* 78.

Heinzel, Hans-Georg. 1988. “Gastric Mill Activity in the Lobster . I . Spontaneous Modes of Chewing.” *Journal of Neurophysiology* 59 (2): 528–50.

Kitano, Hiroaki. 2004 "Biological robustness". *Nature Reviews Genetics*, *5*(11), 826-837.

Knox, Eleanor. 2016. “Abstraction and Its Limits: Finding Space For Novel Explanation.” *Noûs* 50 (1): 41–60.

Prinz, Astrid A, Dirk Bucher, and Eve Marder. 2004. “Similar Network Activity from Disparate Circuit Parameters” 7 (12): 1345–52.

Railton, Peter. 1981. “Probability, Explanation and Information.” *Synthese* 48: 233–56.

Shapiro, Lawrence A. 2000. “Multiple Realizations.” *The Journal of Philosophy* 97 (12): 635–54.

Strevens, M. 2008. *Depth: An Account of Scientific Explanation*. Harvard Univ Pr.

Wilson, M. 1985. “What Is This Thing Called `pain’? -- The Philosophy of Science behind the Contemporary Debate.” *Pacific Philosophical Quarterly* 66: 227–67.