

Causal specificity and the instructive–permissive distinction

Brett Calcott¹

Received: 21 November 2016 / Accepted: 25 February 2017 / Published online: 31 May 2017
© Springer Science+Business Media Dordrecht 2017

Abstract I use some recent formal work on measuring causation to explore a suggestion by James Woodward: that the notion of causal specificity can clarify the distinction in biology between permissive and instructive causes. This distinction arises when a complex developmental process, such as the formation of an entire body part, can be triggered by a simple switch, such as the presence of particular protein. In such cases, the protein is said to merely induce or “permit” the developmental process, whilst the causal “instructions” for guiding that process are already prefigured within the cells. I construct a novel model that expresses in a simple and tractable way the relevant causal structure of biological development and then use a measure of causal specificity to analyse the model. I show that the permissive-instructive distinction cannot be captured by simply contrasting the specificity of two causes as Woodward proposes, and instead introduce an alternative, hierarchical approach to analysing the interaction between two causes. The resulting analysis highlights the importance of focusing on gene regulation, rather than just the coding regions, when analysing the distinctive causal power of genes.

Keywords Causation · Specificity · Instructive · Permissive · Information · Waddington

Introduction

James Woodward has recently outlined three properties of causal relationships that are relevant in many biological contexts (Woodward 2010). *Stability* refers to how robust a causal relationship is across different contexts, *proportionality* captures

✉ Brett Calcott
brett.calcott@gmail.com

¹ Department of Philosophy, Charles Perkins Centre, The University of Sydney, Sydney, Australia

whether the grain of description of cause and effect are appropriately matched, and *specificity* measures the degree to which a cause has fine-grained control over its effects. Each property provides a way to compare causal relationships—a cause can be more stable, more proportional, or more specific than another. These properties:

[...] give us the resources to distinguish among the different roles or relations that causally relevant factors may bear to an outcome (Woodward 2010, p. 89)

Woodward identifies some candidate issues in biology where such resources could be usefully applied. The first of these is the controversial causal parity thesis. The causal parity thesis originates in Developmental Systems Theory (Oyama et al. 2003), and Woodward interprets the thesis as claiming that there is “no principled basis for distinguishing between the causal role(s)” in development (Woodward 2010, p. 289).¹ Woodward suggests that causal specificity may provide such a principle, giving genes a special status because they play a more specific causal role in development than other causes. Causal parity is a topic which has garnered much attention in philosophy, and using causal specificity to identify a special role for genes has been explored and debated in a number of publications (see Waters 2007; Weber 2013, 2016; Griffiths et al. 2015).

The primary aim of this paper is to examine another candidate issue suggested by Woodward which, unlike causal parity, has received no subsequent attention.

[...] the (elusive) contrast between “permissive” and “instructive” causes sometimes employed by biologists might be similarly understood in terms of the idea that permissive causes are generally non-specific in comparison with instructive causes (Woodward 2010, p. 317).

The distinction between instructive and permissive causes arises in a similar context to the causal parity debates, for the heart of the issue is about attributing causal responsibility in development. The distinction is typically applied when talking of induction—the process by which cells in a developing embryo influence the developmental fate of surrounding cells. A key experiment on induction was performed by Spemann and Mangold in 1924, where they showed that a particular cluster of cells—the *organizer*—was responsible for *inducing* an entire body part to form (see De Robertis 2006, for an overview). Spemann and Mangold demonstrated this capacity with a striking experiment, transplanting a cluster of cells into a salamander egg to produce Siamese twins—an embryo with two heads. The initial supposition was that this cluster of cells imparted signals that directed or *instructed* the suite of ensuing developmental processes that formed the additional head. Further experiments, however, revealed that the capacity to perform these processes was already inherent in the cells, and that induction simply released or *permitted* an existing competence to be expressed. The “organizer” turned out to trigger a simple developmental switch, eliciting one of two distinct prefigured developmental pathways.

¹ A reading more consistent with the original thesis is that we should treat causes with “consistency with respect to a criterion” (Oyama 2003, p. 183)—an issue discussed further in Griffiths and Gray (2005). If this is right, then Woodward’s properties provide the very kind of criteria that Developmental Systems Theory would welcome, assuming they are consistently applied.

The literature appears to be divided on whether instructive induction occurs at all. Gilbert (1991) provides a number of cases of induction which he claims are instructive, whilst Kirschner and Gerhart claim that “[...] embryonic induction turned out to be a permissive process”, and go on to say that the alternatives imagined by embryologists where signals are genuinely instructive “have never been found” (Kirschner and Gerhart 2006, pp. 126–127). Even if Kirschner and Gerhart are right, and instructive induction never occurs, the permissive–instructive distinction still marks out an important distinction, for when permissive induction does occur, one implication is that the *instructions must be coming from elsewhere*. It is here that the distinction intersects with the issues about the causal primacy of genes, as it is typically supposed that the instructions are coming from the genome. Scott Gilbert draws on the distinction, telling us that developmental systems theorists (those arguing for causal parity) have:

[...] made the error of not assigning instructive or permissive influences in the interactions. To most of the developmental systems theorists, all the participants are on the same informational level. In this way, the genome is just one other participant, just as are cells or the environment. However, the specificity of the reaction [...] has to come from somewhere, and that is often a property of the genome. Embryology has long recognized that in developmental interactions, there are usually instructive partners and permissive partners (Holtzer 1968). Instructive partners provide specificity to the reaction, whereas permissive partners are necessary, but do not provide specificity.

Gilbert links the notion of specificity in biology to instructive and permissive interactions, and asserts that the specific or instructive “partners” will often come from the genome. Accordingly, what primes the cell, giving it the competence to respond to the non-specific, permissive signal is something instructive *in the genome*. If the specificity Gilbert has in mind is adequately captured by Woodward’s *causal specificity*, then perhaps the distinction between permissive and instructive causes can be collapsed onto the existing debates about causal specificity and causal parity.

I shall resist this collapse and instead argue that a closer examination of the structure of permissive and instructive causes exposes two assumptions in the current application of causal specificity to debates about the primacy of genes. First, attempts to “distinguish among the different roles or relations that causally relevant factors may bear to an outcome” (Woodward 2010, p. 89) have focused solely on a competition between the specificities of interacting causes, where bigger is better. I outline an alternative, hierarchical, way of analysing the interaction of causes. Second, it has been consistently assumed that specificity *is for* the coding of proteins, or more broadly, particular products in development. Instead, I show that specificity can be for the conditional response of the genome, thus forging a connection between these causal properties and the regulatory, rather than coding, regions of DNA. A second aim in this paper, then, is to expand the way that these causal properties are used to analyse and clarify the complex interplay of different causal roles in development and evolution.

My strategy in the rest of the paper relies on two key tools. The first is the information theoretic approach to causal specificity outlined in Griffiths et al.

(2015) and elaborated in Pocheville et al. (2017). I introduce these tools in “[Tuning Woodward’s radio](#)” section and rely on them throughout the paper as a check on our intuitions about causal relations. The second tool is the introduction of a model of causal interaction to capture our intuitions about permissive and instructive causes and to direct the formal analysis. The introduction of this model plays a key role in my arguments for, despite the wealth of complex biological details in the recent discussion about causal specificity and causal parity (Weber 2013, 2016; Griffiths et al. 2015), the examples used to reason about contrasting specificity all exhibit a relatively simple causal structure, exemplified by Woodward’s example of a radio. A novel model, I suggest, can direct our thinking in some useful new directions.

Tuning Woodward’s radio

I begin by outlining a simple example—Woodward’s Radio—that exemplifies the current approach to comparing two causes for the same effect. In this example, the on–off switch of a radio is contrasted with its tuning dial. The tuning dial has fine-grained control over what we hear as we can tune to many stations, whereas the on–off switch is relatively coarse-grained; it is causally relevant to whether we hear anything, but has little control over what station we receive. The example captures the idea that, although the switch and the dial are both causes of what we hear, the dial is a *more specific cause*. Woodward introduced the example to highlight a similar causal structure in biology: Contrasting the switch-like role that RNA polymerase plays in protein synthesis to the fine-grained role that DNA plays [see Waters (2007) and Griffiths et al. (2015) for discussion of this example].

We can sketch this claim about the radio in more detail. In the interventionist framework, causes and effects are variables. To capture the radio, we need two cause variables, S for the on–off switch and D for the tuning dial, and an effect variable, A , for the “audio” we hear. The on–off switch clearly has just two states. I shall assume that there are only eight dial settings, each of which tunes to a different station. Our effect variable thus has nine states—one each for whatever is playing on the radio stations and one for silence. All possible states of the system and their relationships are summarised in Table 1. The combination of causes (the switch and the dial) make up the rows, and the different effects (what we hear) are in the columns. An \times marks the spot in a column, telling you the result of setting one of the sixteen possible dial/switch combinations.

Representing the radio as I have in Table 1 reveals something important that Woodward only mentions in passing (Woodward 2010, p. 307). The fine-grained control of the tuning dial only exists when the radio is on (top half of the table). When the radio is off, the tuning dial fails to be a cause at all (bottom half of the table). In contrast, the on–off switch is always a cause, no matter what channel is chosen. How can we compare a cause that is fine-grained some of the time (the dial) with a cause that is not fine-grained all of the time (the switch)? We can use a formal approach to shed light on this question.

Table 1 Capturing Woodward’s radio: an on–off switch, S , and a tuning dial with eight states, D , together control what we hear on the radio, A

		$A = a_1$	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_{silence}
$S = s_{\text{ON}}$	$D = d_1$	×								
	d_2		×							
	d_3			×						
	d_4				×					
	d_5					×				
	d_6						×			
	d_7							×		
	d_8								×	
$S = s_{\text{OFF}}$	$D = d_1$									×
	d_2									×
	d_3									×
	d_4									×
	d_5									×
	d_6									×
	d_7									×
	d_8									×

A formal approach

Woodward proposed that the upper limit of fine-grained influence is a one-to-one (bijective) mapping between the values of the cause and effect variables: every value of an effect variable is produced by one and only one value of a cause variable and vice versa (Woodward 2010). Griffiths et al. (2015) showed that this idea can be generalised to the whole range of more or less specific relationships using an information-theoretic framework. In their approach, causal specificity can be measured by the causal mutual information between the cause variable and the effect variable. This formalises the idea that, other things being equal, the more a cause specifies a given effect, the more knowing how we have intervened on the cause variable will inform us about the value of the effect variable. Though a mutual information measure is symmetrical, this symmetry will typically not hold if the underlying probabilities are generated by a causal model.

A simple way to apply this measure to the radio is to convert the preceding table into a joint probability distribution (see Table 2). To generate this distribution I need to assign probabilities to my choice of dial settings, and to how often I turn the radio on. One way to do this is to presume we are performing an ideal experiment, and assign a uniform distribution over the causes. Thus, each possible combination of causes has a probability of $1/2 \times 1/8 = 1/16$. Note that the probabilities in Table 2 are not simply the observed co-occurrences of the settings of the dials and switches and what we hear. That is, the entries in the table are not simply $P(s_i, d_i, a_j)$. Rather, each probability in the table tells us how likely it is that I hear a particular station (or silence) when I set, or *manipulate*, the dial and switch to

Table 2 Woodward’s radio, captured using a joint probability distribution generated *under intervention*

		$A = a_1$	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_{silence}
$\widehat{S} = s_{\text{ON}}$	$\widehat{D} = d_1$	1/16								
	d_2		1/16							
	d_3			1/16						
	d_4				1/16					
	d_5					1/16				
	d_6						1/16			
	d_7							1/16		
	d_8								1/16	
$\widehat{S} = s_{\text{OFF}}$	$\widehat{D} = d_1$									1/16
	d_2									1/16
	d_3									1/16
	d_4									1/16
	d_5									1/16
	d_6									1/16
	d_7									1/16
	d_8									1/16

Generating this table requires choosing a probability distribution over each of the causes. Here, I have chosen a uniform distribution over each cause

particular values, multiplied by the probability that I choose that particular combination of dial and switch. We write this as $P(\widehat{s}_i, \widehat{a}_j)P(a_k|\widehat{s}_i, \widehat{a}_j)$, where the hat signifies an intervention (flicking a switch or setting a dial). So this is a joint probability distribution generated under intervention, or an *interventional distribution*.

The two values we want to compare are the causal specificity of the switch, S , and the specificity of the dial, D , for the effect A . This is $I(\widehat{S}; A)$ and $I(\widehat{D}; A)$. The distribution in Table 2 shows us what happens under all combinations of the two causes. To calculate the specificity of one of these causes, we need to summarise the probabilities over the other, non-focal, cause. Table 3 shows the resulting distributions produced for both the switch and the dial when this is done. These are the interventional distributions for each separate cause. To calculate Griffiths et al.’s measure of specificity, we simply calculate the mutual information between the cause and effect variable in these interventional distributions. We can do this because the mutual information, calculated on a distribution generated under intervention, is equivalent to the causal specificity measure. In general, to calculate any information theoretic measure (such as specificity) relating a cause and an effect, we generate a joint interventional distribution using a causal model, and then use the standard information theoretic measures on this modified distribution.

Table 3 Two distinct interventional distributions, both constructed from the full distribution in Table 2 by summing over the background cause

	$A = a_1$	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_{silence}
<i>Summed over all values of \widehat{S}</i>									
$\widehat{D} = d_1$	1/16								1/16
d_2		1/16							1/16
d_3			1/16						1/16
d_4				1/16					1/16
d_5					1/16				1/16
d_6						1/16			1/16
d_7							1/16		1/16
d_8								1/16	1/16
<i>Summed over all values of \widehat{D}</i>									
$\widehat{S} = s_{ON}$	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	
s_{OFF}									1/2

The mutual information in these resulting interventional distributions provides the specificity of the focal cause. On the top, the specificity of the dial is $I(\widehat{D}; A) = 1.5$, and on the bottom, the specificity of the switch is $I(\widehat{S}; A) = 1.0$

Given the probabilities in Table 3, the specificity of the switch is 1 bit and the specificity of the dial is 1.5 bits.² The dial, as Woodward suggested, is more specific than the on–off switch, though only by a small amount in this case. In fact, if we only had four radio stations, the specificity of the two variables is equal. In contrast, if we increase the number of dial settings and the number of radio stations (maintaining bijectivity), then the specificity of the tuning dial increases linearly: with 16 settings, the specificity is 2.0 bits, and with 32, it becomes 2.5 bits, and so on. The specificity of the on–off switch, however, remains fixed at 1.0 bits in all cases. It cannot get larger, because the maximum specificity a simple switch can have is 1 bit.³ So applying the formal apparatus confirms our intuitions, and also

² Here is a simple approach to computing causal specificity. A joint probability distribution over two discrete random variables X and Y can be represented using a table (such as those in Table 3), where each cell contains $P(x, y)$ —the probability of observing both $X = x$ and $Y = y$. To calculate the (non-causal) mutual information between the two variables, we first calculate the pointwise mutual information for each cell, $\log \frac{P(x, y)}{P(x)P(y)}$, noting that numerator here is the value of the cell, and the denominator is simply the product of the sums of both the row and column in which the cell lies. By convention, if either $P(x)$ or $P(y)$ is zero then pointwise mutual information is 0. The mutual information is then the expectation of the pointwise mutual information computed in all cells, thus $I(X; Y) = \sum P(x, y) \log \frac{P(x, y)}{P(x)P(y)}$. Now, if the joint distribution in the table was produced under intervention, then this same operation on the table computes a causal specificity measure. To see why, note that pointwise value we now calculate will have a numerator (the cell value) of $P(\widehat{x})P(y|\widehat{x})$ and a denominator of $P(\widehat{x})P(y)$. Cancelling $P(\widehat{x})$, we now compute $\log \frac{P(y|\widehat{x})}{P(y)}$ for each cell, and the expectation of this is $I(\widehat{X}; Y) = \sum P(\widehat{x})P(y|\widehat{x}) \log \frac{P(y|\widehat{x})}{P(y)}$, which is causal specificity as defined by Griffiths et al.

³ This is because the mutual information between two variables cannot exceed the lowest entropy of the two variables, and the maximum entropy of a variable with two states is 1 bit. Formally, $I(X; Y) \leq \min(H(X), H(Y))$.

makes explicit some assumptions we have made in the process—a uniform distribution over causes, and the number of radio stations required to ensure the dial is more specific.

Hierarchical analyses of interacting causes

The analysis of Woodward's radio in the last section reflects the way causal specificity is currently being applied to the debates about causal parity. In the context of Woodward's radio, the aim is to answer this question:

- (C) How can we characterize what is distinctive about the causal relationship between (a) the tuning dial and what we hear, in contrast to (b) the on–off switch and what we hear.

If we decide that specificity captures the characteristics of interest, then this question becomes: Which cause has more fine-grained control over what we hear, the on–off switch or the tuning dial? As we saw above, once we had a causal model and had fixed some probabilities, we could give a definitive answer to this question using information theoretic measures of causation. This approach to analysing the interacting causal relationships directly contrasts the overall contribution of two causes to the same effect. Let us call this a *competitive* approach to analysing how the two causes interact—we calculate their specificity, and the bigger one wins.

An alternative approach is available, however, which treats the relationship between the causes as *hierarchical*. To perform this analysis within the context of Woodward's radio, we instead ask the following two questions:

- (H1) How can we characterize what is distinctive about the causal relationship between (a) the tuning dial and what we hear when the radio is on, in contrast to (b) the tuning dial and what we hear when the radio is off?
- (H2) How can we characterize the causal relationship between the on–off switch and the effect it has on altering the causal relationship between (a) and (b) described in (H1)?

This approach is hierarchical because we assume a *background cause* (the switch) controls the causal relationship between a *foreground cause* (the dial) and an *effect* (what we hear). The aim in this analysis is to contrast the foreground causal relationships produced across different backgrounds, and then to characterise how the background controls these changes. Let me sketch briefly how this kind of analysis might go, and contrast it to what I did in the previous section.

First, we set the radio on and measure how specific the tuning dial is for what we hear. Formally, we measure $I(\widehat{D}; A | \widehat{S} = s_{\text{ON}})$. This can be calculated using just the set of relationships in the top half of the Table 1. As discussed above, we still need to provide a causal distribution for the states of the dial. If we assume they are uniformly distributed, then the dial will have 3 bits of specificity for what we hear.

Similarly, we can calculate how specific the dial is for what we hear when the radio is off, obtaining a value of 0 bits.⁴

What about the specificity of the on–off switch? The assumption made in the question (C) was that the specificity of the switch, like that of the dial, is *for what we hear*. But (H2) asks something different. Flipping the switch changes the way the *tuning dial maps to what we hear*. When the radio is off, each setting of the dial maps to silence. When the radio is on, each setting of the dial maps to a distinct radio station. So, rather than measuring the specificity the switch has for the stations, we measure the specificity the switch has for the distinctive ways that the dial maps to the stations. The switch has two states, on and off, and each state produces a distinct mapping. Assuming a uniform distribution over the causal variable, the switch has 1 bit of specificity for these mappings.

We now have some tentative answers to the two questions posed in this alternative analysis. What is distinctive about the tuning dial when the radio is on is that it is causally specific for what we hear (3 bits of specificity). When it is off, this specific relationship is not present (0 bits of specificity). The specificity that the switch has for this mapping relationship is 1 bit. As before, the formal analysis quantifies the amount of specificity in both causes, and thus we can say that the tuning dial *is* a more specific cause than the switch. This claim comes with two caveats though. First, the dial is only more specific for what we hear when we fix the radio on. Second, the two specificities we have measured are not for the same effect. Whilst the specificity of the dial is for the radio stations, the specificity of the switch (1 bit) is what permits us to change the specificity of the tuning dial (from 0 to 3 bits).

Does this alternative analysis deliver a useful understanding of the causal relationships in the radio? It certainly provides one way to interpret the causal relationships in a radio. The tuning dial is useful because it gives us fine-grained control over the various stations *when the radio is on*. Our ability to manipulate the on–off switch, on the other hand, is what puts us in the position to exploit this fine-grained relationship, by controlling the background conditions which make this fine-grained control possible.⁵

So we can make sense of the hierarchical approach, but do we have reason to prefer this analysis over the competitive approach? Importantly, if we do use this approach, what reasons have we for thinking of the switch as the background cause and the dial as the foreground cause, and not the other way around? The structure of the causal graph that represents these relationships is symmetric, so if we are to decide, we need to say why we should interpret the model in this particular way,

⁴ A specificity of zero tells us that the dial is not a cause at all. So I could simply say that the dial is a cause when the switch is on, but not a cause when the switch is off, at least in this case. Having a graded measure of causation will, however, turn out to be important in more complex cases in the following sections.

⁵ Woodward’s discussion appears to shift between these two modes of analysis. He tells us that “[...] the dial gives one relatively fine grained control over which station is received, *assuming that the switch is on*.” (Woodward 2010, my italics), thus drawing attention to specificity against one background. When he talks of the switch, however, he appears to consider it against all backgrounds: “one can’t modulate or fine-tune which station is [received] by varying the state of the switch.”

treating the causes differently. One possible response is that the order we do things identifies which cause is treated as the background cause. First, we turn the radio on, thus setting up the background conditions against which we tune it. Maybe this is how the radio is used; but the contrary is also possible: if I know the station I want, then I can tune it first, then turn it on. The radio example simply runs out of steam here—we need something more complex and more biologically grounded to justify when this analysis is more appropriate. I introduce such a model in the next section, and then return to this issue in “[Hierarchical analyses and different dimensions of manipulation](#)” section.

Building a Waddington box

Woodward’s radio gave us a simple tool for thinking through how specificity can be used to compare different causes for the same effect. It also provided the basis for constructing a model in which we could formally measure this specificity, and performing this analysis helped clarify the example and its underlying assumptions. My aim in this section is to sketch a new thinking tool that will serve the same role as the radio, but one that is suited to capturing the permissive–instructive distinction. I shall spend some time introducing the example for, as we shall see, the example itself can influence what we think of as the most appropriate way to analyse a set of interacting causal relationships.

To be useful, the example must resemble the biological facts of interest, so I shall build on a well-known biological metaphor for development: Waddington’s epigenetic landscape. Waddington’s pictures depict marbles rolling down a landscape of hills and valleys, eventually coming to rest at the lowest points (Fig. 1a). The valleys are developmental pathways and the resting points represent cell fates, or tissue types. Waddington imagined the shape of landscape was influenced by various guy ropes that were pegged down, and thought of these pegs and ropes as controlled by the genes. Although Waddington’s ideas lost prevalence once Jacob and Monod’s operon model was established (Gilbert 1991), his landscapes still figure prominently as a tool for thinking about developmental processes, and are now reinterpreted as representing the dynamics produced by a gene regulatory network (Wang et al. 2010, 2011; Wilkins 2008; Siegal and Bergman 2002; Huang 2012; Bhattacharya et al. 2011). Thus, the metaphor continues to guide our thinking about the dynamic interaction of causes in development.

I shall join Waddington’s metaphor to another well-known idea: Galton’s box, otherwise known as the bean machine or quincunx (see Fig. 1b).⁶ The device was invented by Francis Galton to provide a clear and visually striking way of generating a normal probability distribution. It consists of a vertical board with a series of interleaved pins, and a set of buckets at the bottom. Metal balls are introduced from a single central point above the pins (Galton used lead shot). They rebound on the pins eventually collecting in the buckets at the bottom generating a

⁶ My thanks to Paul Griffiths for drawing my attention to the Galton box.

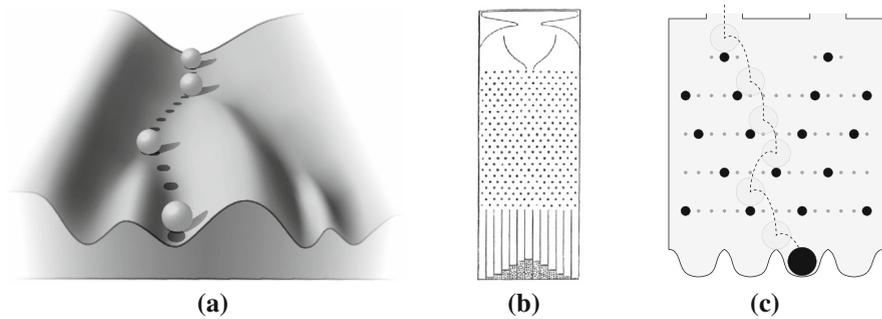


Fig. 1 **a** Waddington's epigenetic landscape: a visual metaphor (modified from https://figshare.com/articles/_Waddington_s_8220_Epigenetic_Landscape_8221_/620879, Creative Commons license CC BY 4.0). **b** The Galton box: a device for visually generating a normal distribution (modified from [https://commons.wikimedia.org/wiki/File:Quincunx_\(Galton_Box\)_-_Galton_1889_diagram.png](https://commons.wikimedia.org/wiki/File:Quincunx_(Galton_Box)_-_Galton_1889_diagram.png), Public domain image). **c** A Waddington box, showing one possible configuration of the pins, and one path the ball may take if it is placed in the *left* slot

visible distribution, with most in the center and fewer toward the edges. Galton used the device to demonstrate the central limit theorem (Bulmer 2003, 183).

Putting these ideas together gives us a *Waddington Box*: a device which, like Galton's box, generates probability distributions over buckets. This device, however, has a number of ways to manipulate the resulting probability distribution. First, I can decide on the layout of the pins: adding, removing, and relocating them. Different pin layouts will affect the way that the ball bounces around and the probability that it comes to rest in any particular bucket. Second, the box has multiple slots on top where I can place the ball, rather than the single slot of Galton's box. Each slot drops the ball onto the pins in a different place, and thus also affects which bucket the ball ends up in (see Fig. 1c).

A Waddington box is a mechanical version of Waddington's metaphor. The arrangement of the pins, like Waddington's pegs and guy ropes attached to the epigenetic landscape, influences the path that the ball can take. The slot where the ball is placed is the initial starting place on the landscape and the final bucket where the ball comes to rest is the resulting state of the system: a cell or tissue type. The addition of different starting points on the surface (the slots on top) also fits with Waddington's own view, who considered induction an initial push on a landscape shaped by genes (Gilbert 1991, p. 140), and of development as series of binary branching points defining alternative pathways (Wilkins 2002, pp. 105–108).

Like Woodward's radio, the Waddington box has two causes that interact to produce a single effect. Both the slots and the layout of the pins are causes of the ball's destination, according to Woodward's undemanding definition of causation, "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 or the probability distribution of Y would change (Woodward 2010, p. 290).

Let P denote the different pin layouts, S the slots where we put the ball, and B the buckets where the ball ends up. If we treat the layout of the pins, P as the background circumstances, then in many cases (though not all, as we shall see below) if we intervene on S , then the distribution over B would change. Similarly, if we treat the slots, S as a background circumstance, then intervening on the pin layouts, P , can likewise change the distribution over B .

The pins and the slots are both causes, but, importantly for our task here, the differences between them plausibly captures the contrast between instructive and permissive causes. We can think of the slots in which we place the ball as merely inducing an existing response that is prefigured in the current layout of the pins. A first pass, following Woodward's idea, might be that the "mere induction" of the slots is due to the limited range of choices—there are, after all, just two slots. In contrast, there is a large range of different pin layouts we could choose, and this choice of layouts allows us to fine tune the path and final destination of the ball coming from each of the slots. This contrast certainly looks like it may be characterised formally in terms of causal specificity.

As with the radio, a formal approach can help clarify and confirm (or not) these intuitions. To apply the measures, however, we first need a concrete model. The Waddington box depicted in Fig. 1c has two slots, $\{s_1, s_2\}$, and four buckets, $\{b_1, b_2, b_3, b_4\}$, but how many pin layouts does it have? The number could be enormous. There are 66 holes in which pins can be placed, so if each hole can either be empty or contain a pin then the total number of distinct pin layouts is $2^{66} = 73,786,976,294,838,206,464$. By setting some simple limitations on the layouts—such as ensuring there are spaces between pins—we can reduce this number to a more manageable 3,306,664 layouts (see "The layouts" section in "Appendix" for details). There is a second problem: how do we determine the distributions these different layouts produce—neither building all these boxes or realistically simulating the physics of such devices is feasible. Again, we can simplify things by defining a simple rule-based physics governing how the ball interacts with the pins as it falls through the box (see "The physics" section in "Appendix" for details). This rule-based approach still maintains the key aspects of such a box—they are sufficient to simulate Galton's box, for example. Using this approach, I used a computer simulation to calculate the probability of all paths a ball can take from a slot to a bucket and generate a distribution over the buckets for both slots in each of the 3,306,664 layouts.

Analysing the causes in the box

We now have a model we can analyse, but what do we want to measure? There are three key intuitions about the permissive–instructive distinction that are reflected in the Waddington box:

- (W1) The choice of slots induces a response in the buckets (the destination of the ball).

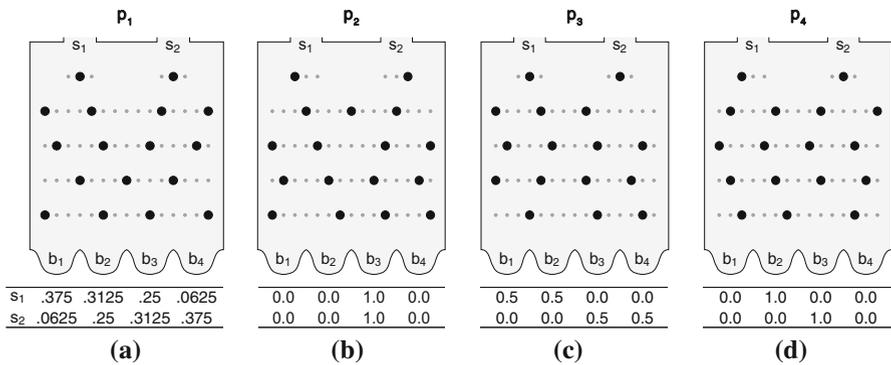


Fig. 2 Four of the 3,306,664 possible layouts for the Waddington box. Below each box is the resulting probability distribution over the buckets for each slot

- (W2) Simply citing the slots as a cause, however, appears to miss something crucial.
- (W3) The crucial missing piece is explained by adverting to the layout of the pins.

My aim is to characterise these intuitions in a principled way drawing only on the measurable properties of the causal relationships in the box. I show that these intuitions call for a hierarchical analysis rather than a competitive one. That is, these intuitions are best captured by answering the two questions I outlined earlier: determining what is distinctive about the causal relationship between the slots and the buckets, and then determining how the pins can be manipulated to control this causal relationship.

To make these arguments, it will be useful to refer to some particular layouts in the Waddington box. Figure 2 shows four pin layouts—four settings of the P variable—and the resulting distributions over the buckets they generate when placing the ball in each of the slots. In each case, I assume that we place the ball in each slot with equal probability. I describe them below:

- p_1 : The pins are evenly spaced across the board. The result is two opposingly skewed distributions over the buckets.
- p_2 : The pins are arranged to create a funnel. Both slots send the ball into bucket b_3 .
- p_3 : The pins evenly divide each slot between two of the buckets. Slot s_1 goes to buckets $\{b_1, b_2\}$; slot s_2 goes to $\{b_3, b_4\}$.
- p_4 : The pins are arranged to funnel each slot into a single bucket. Slots s_1 goes to b_2 , and slot s_2 goes to b_3 .

What is a permissive cause?

In the hierarchical analysis I sketched in “[Tuning Woodward’s radio](#)” section, the first step was to outline what is distinctive about the foreground cause given certain

backgrounds. With respect to the Waddington box, this amounts to identifying what is distinctive about the causal relationship between the slots and the buckets given certain pin layouts. Recall that the origin of the permissive–instructive distinction was the induction of a cell or tissue type in development. A permissive cause thus plays the role of a reliable developmental switch: when it is on, one cell type or tissue develops; when it is off, a different cell type or tissue develops. Of the four examples given in Fig. 2, only LAYOUT p_4 produces a reliable switch, mapping each slot to one and only one bucket. This suggests that, rather than thinking of the slots as a permissive cause in general, we should think of them as permissive only in certain layouts. Our first goal, then, is to state clearly what is distinctive about the LAYOUT p_4 in contrast to the other layouts.

A first guess is to simply contrast the specificities that the slots have in each of these layouts, in the same way that I contrasted the specificity of the tuning dial when the radio was on and off. If we measure the specificity of the slots in each of these layouts, we find that the specificity of LAYOUT p_1 is around 0.184, the specificity of LAYOUT p_2 is 0.0, and both LAYOUT p_3 and LAYOUT p_4 have a specificity of 1.0.

Measuring the specificity across these layouts tells us two things. First, low specificity itself is not the mark of a permissive cause. In LAYOUT p_2 the slots have zero specificity because the slots are *not a cause at all*—where we place the ball has no effect on the resulting bucket. LAYOUT p_1 has the next lowest specificity—it clearly changes the distribution over the buckets producing the mirror image, but this kind of control is not switch-like. I draw attention to this because both Woodward’s suggestion that “permissive causes are generally non-specific in comparison with instructive causes” might be interpreted as meaning permissive causes are simply low specificity causes. These examples show that low specificity by itself cannot be sufficient to capture a permissive cause, for low specificity can be indicative of both lack of any causal control or a noisy relationship between cause and effect.

Second, notice that LAYOUT p_3 has precisely the same specificity as LAYOUT p_4 . LAYOUT p_3 acts like a fuzzy switch, rather than a reliable one: you can control which subset of buckets a ball goes into, $\{b_1, b_2\}$ or $\{b_3, b_4\}$. LAYOUT p_4 , in contrast, is a precise switch—placing a ball in either slot tells us precisely which bucket the ball will fall into. Given that the specificity measure fails to distinguish these two layouts, it follows that specificity by itself is insufficient to capture what is distinctive about a permissive cause. What, then, is distinctive about the layout LAYOUT p_4 ?

Let us look at the difference between LAYOUT p_3 and LAYOUT p_4 . Notice that, in LAYOUT p_4 only two buckets are ever used, rather than four in LAYOUT p_3 . So although the possible range of effects is four buckets, this particular layout limits the actual range of buckets to the subset $\{b_2, b_3\}$. We can capture this reduction in buckets by looking at the entropy of the effect variable, $H(B|P = p_i)$, where p_i is one of the four layouts. For LAYOUT p_3 this value is 2.0 bits; for LAYOUT p_4 this value is 1.0 bit. We can see that, in LAYOUT p_4 , the entropy of the effect is equal to the specificity:

$$H(B|P = p_4) = I(\widehat{S}; B|P = p_4) \quad (1)$$

When the specificity of a cause for an effect is equal to the entropy of that effect, then we can say that a cause *precisely controls* its effect. That is, if we know the state of the cause, there is nothing more we need to know to determine precisely what the effect is.⁷

What is distinctive about LAYOUT p_4 , then, is that the slots precisely determine the buckets, and can only do this because the number of actual buckets used in this layout has decreased. For if a cause is to precisely determine its effect, it must have at least as many states as that effect.

The notion of a cause *precisely* determining its effects is not captured by any of Woodward’s three properties, specificity, proportionality, or stability, though it is clearly an important characteristic of a causal relationship. One reason for its absence may be that a bijection—Woodward’s preferred way of introducing specificity—maximises specificity and is also perfectly precise. But these two properties can easily come apart, as we see in the contrast between LAYOUT p_4 and LAYOUT p_3 . To exaggerate what is going on here, consider a radio dial which has 100 settings, but each of these settings randomly tunes into one of two distinct radio stations (thus, there are 200 possible radio stations). The dial on this radio with its 100 dial settings is far more causally specific than the dial with only four settings that we looked at in “[Tuning Woodward’s radio](#)” section. But it is less precise, for despite the increase in fine-grained control, the dial never fully determines the station.

The notion of precision is also a plausible interpretation of Waddington’s concept of canalisation:

The main thesis is that developmental reactions, as they occur in organisms submitted to natural selection, are in general canalized. That is to say, they are adjusted so as to bring about one definite end-result regardless of minor variations in conditions during the course of the reaction (Waddington 1942).

Precision, as I’ve defined it above, is about obtaining “one definite end-result” given the layout and slot are fixed. A close examination of LAYOUT p_4 shows that a ball can take a number of different paths from s_2 to b_2 . These count as “minor variations in conditions during the course of the reaction”, and despite this, a single end-result is obtained. Finally, as I explain in detail below, this canalised capacity is possible because these reactions can be “adjusted” by manipulating the layouts. If this is right, then we might call a permissive cause a *canalised switch*.

I can now address the intuition that the slots merely induce a response, and that this response is already prefigured in the layout of the pins. The idea that a permissive cause needs some additional explanation in terms of a second instructive cause can be explained informally in the following way:

1. The slots have a mere two states.

⁷ We can state this equivalently by saying that the conditional entropy is zero: $H(B|S, P = p_4) = 0$. But the formulation in Eq. (1) will serve us better below.

2. The buckets, however, have four states. That is, the number of possible states exceeds that which can be precisely controlled by slots.
3. Despite this, in some background conditions, the slots do precisely determine the state of the buckets.
4. This precise control is only possible because the background conditions have ensured that:
 - (a) the actual number of buckets used has been reduced, and
 - (b) each state of the slots maps directly to one and only one of these buckets.

We can state the same thing formally. If we define $H_{\max}(x) \equiv \log_2(x)$, then we have $H_{\max}(S) < H_{\max}(B)$. In certain settings of P , it is possible to both (a) reduce $H(B)$, and (b) increase $I(\hat{S}; B)$ so that they are equal.

In short, a permissive cause has limited causal capacity but nevertheless exhibits precise control over an effect whose range of possible states exceeds this capacity. Because this only occurs under some specific background conditions, we look to those conditions for an explanation. That is, we seek an instructive cause.

The fact that this distinctive causal character—the precision—only occurs in some background conditions is crucial. To see why, notice that the distinctive character of LAYOUT p_4 I described above is very close to Woodward’s notion of a “pure switch” that he uses to describe the on–off switch in the radio (and by implication, the RNA polymerase in protein synthesis) (Woodward 2010, pp. 307–307). That is, the radio on–off switch, like the slots in LAYOUT p_4 , precisely determine their effects. Once the dial is fixed, the state of the switch tells us precisely what we’ll hear—a specific station or silence. And, as Woodward points out, the switch only allows access to a limited range of the effects—only the station currently tuned by the dial and silence can be chosen. There is, however, a key difference between the radio’s on–off switch and LAYOUT p_4 . In the radio, manipulating the other cause, the tuning dial, has no effect on the causal relationship between the switch and what we hear. The on–off switch in a radio is always a switch, no matter the background conditions.⁸ In the Waddington box, in contrast, the distinctive causal relationship of a precise switch is only produced in some background conditions. What makes the slots a permissive cause is not just that they are precise, but the fact that this causal relationship could have been different. These differences are brought about by manipulating the pin layouts, so the pin layouts are a difference-maker with respect to the causal relationship between the slots and the buckets. This brings us to the second part of the hierarchical approach: characterising the control that the pin layouts (the background cause) have over the causal relationship between the slots and the buckets (the foreground cause).

⁸ To avoid any confusion: I’m now treating the on–off switch as a foreground cause and the tuning dial as a background cause. This is the opposite of what I did in “[Tuning Woodward’s radio](#)” section. This is no cause for alarm, I am making a different point here.

What is an instructive cause?

What makes the pins an instructive cause? The intuition was that the pin layouts allowed us to fine tune the relationship between the slots and the buckets. To establish this, we need to look at the causal relationship between the layouts and the mappings between the slots and buckets.

Each of the 3,306,664 possible layouts produces a mapping from slots to buckets. A mapping consists of two distributions over the buckets, one for each of the slots. These two distributions are what can be seen below the diagrams of the four Waddington box configurations in Fig. 2. Each distribution captures the probabilities of a ball ending up in each of the buckets given it is placed in that particular slot. A mapping thus defines the conditional response of a particular pin layout to the two possible inputs from the slots. With respect to the biological inspiration for this model, the idea of a mapping captures the plasticity of a developmental response, or the divergence of two developmental pathways given some prior input.

If we consider two mappings distinct when either one of these distributions is different, then we can represent the mappings as a variable, M , with as many states as there are distinct mappings across all pin layouts. Enumerating all the mappings of the 3,306,664 layouts, we find a total of 5683 distinct mappings. So there is not a simple bijective relationship between layouts and mappings. For example, there are 19,426 different pin layouts that produce precisely the same distribution seen in LAYOUT p_4 and 232 different layouts that produce the same distribution as seen in LAYOUT p_2 .

Clearly, however, in many cases, manipulating the pins, P , will change the mapping, M . Thus, the layouts are causes of the mappings, assuming an interventionist account of causation. We can thus measure the causal specificity of the relationship between the layouts and the mappings in our model. Assuming a uniform distribution over the layouts, the resulting specificity of the pins for the mappings, $I(\hat{P}; M)$ is approximately 8.2 bits. How specific is that? For the radio tuning dial in Woodward’s radio to have the equivalent specificity it would need more than 256 positions that tuned to 256 unique stations. Despite the fact that there is not a bijective mapping, I think this counts as fine-grained control. The pins are a highly specific cause of the relationship between slots and buckets.

This size of specificity can be highlighted by contrasting it with the results from a competitive approach—the kind of analysis I first used to analyse Woodward’s radio in “[Tuning Woodward’s radio](#)” section. Recall that to compare these two causes fairly we must assess the specificity of each causal variable by summing the probabilities across all states of the other cause. We have all the data from the simulations. What answers do we get? The specificity of the pin layouts for the buckets, $I(\hat{P}; B)$ amounts to a paltry ≈ 0.26 bits. This value is, in fact, less than the specificity that the slots have for the buckets: $I(\hat{S}; B) \approx 0.67$ bits, despite the fact that there are over three million layouts and only two slots.

Notice that even without performing any measurements, we know that the specificity of the pin layouts could not exceed 2 bits, for it is limited by maximum entropy of the buckets. Increasing the number of buckets, then, would be one way to

raise the specificity of the layouts for the buckets. So there are ways we could feasibly elevate this value. Such a pursuit might be successful, but it should not divert us from the lessons learnt above. First, that we can make sense of the idea that the specificity of the pin layouts reflects our ability to control a set of distinctive pathways that a ball can take from each slot, rather than modify the resulting distribution produced from both slots. Second, that this specificity is not limited by the number of states of the buckets, for the effect of interest is actually a mapping between the slots and buckets.

The aim of this section was to formalise the intuitions about permissive and instructive causes within the Waddington box. Using a hierarchical analysis of the interacting causes, it is possible to ground these intuitions using formal tools. I have shown that the results are not simply a matter of comparing the specificity of two causes. The slots are permissive causes when they are precise switches for the buckets, and the pins are an instructive cause because they provide fine-grained control over the conditions that produce this distinctive causal relationship. If this is right, then the relationship between permissive and instructive is not one of causal selection: where one cause is identified as more important because it is more specific for the same effect (see, for example Weber 2013). Rather, it is one of explanation: the specificity that the pins have for the mappings explains how it is possible to manipulate the causal relationship between slots and the buckets, and thus generate a precise switch—one that can be controlled by a simple cause.

Hierarchical analyses and different dimensions of manipulation

In the last section, I showed that we can make sense of, and measure, the specificity that one cause has over another causal relationship. I argued that this was the best way to capture our intuitions about the pin layouts in the Waddington box having fine-grained control. This reflected both the massive number of different states they could assume, and the clear impact this had on manipulating the pathways the balls could take. This distinction relied on a particular approach to analysing the causes that I called hierarchical, and I contrasted it with the a familiar competitive approach. In “[Tuning Woodward’s radio](#)”, however, I questioned what basis there was for treating one cause as a background cause and another as foreground. It was difficult to make this argument in the context of the radio, but the biological inspiration for the Waddington box gives us some reasons for demarcating background and foreground causes.

Here is a natural way to think of how we might explore the causes in a Waddington box. First, we decide on a layout of the pins. Then, we drop a number of balls into both slots, looking at the pathways they take, and the distributions they generate.⁹ So first we fix the layout, then manipulate the slots. It is possible to do it the other way around, of course. That is, we could fix the slot we use (say, just slot s_1), drop some balls down it, generate a distribution, change the layout, drop some more balls, and so on. But, as we saw in the last section, a central concern was how

⁹ This is, in fact, how the simulations that generate the distributions actually work.

the manipulation of the layout controlled the pathways that lead from *both slots*, not just one. If we reflect on what the causes in the Waddington box represent, it becomes clear why we want to look at both pathways simultaneously. The slots represent some inductive chemical, and (at least in the case of differentiation in development) the full range of these different states (on and off) will actually occur during development in different cells. In contrast, the pins represent a genetic state that will remain fixed during development, and only change across an evolutionary time scale. Thus, although both pins and slots are both causes, it is natural to interpret them as being manipulated in quite different ways with respect to one another. We are interested in exploring the full range of manipulation available for one of causes (the slots) whilst fixing the state of the other cause (the pins). The contrary is not the case.

We now have an answer to why the hierarchical analysis is done this way. Both the layouts and the slots are difference-makers, but the way we manipulate them with respect to one another is not symmetrical. Our interest lies in the distributions produced by both slots once a layout is fixed. And it is natural to think of the layouts as a fixed background cause because, in our biological analogy, this cause does actually remain fixed during development, whilst the other causes actually differ. We can think of these as two different dimensions of manipulation—one occurring at a developmental time scale, the other at an evolutionary time scale.

Causal specificity and gene regulation

The permissive–instructive distinction is used to describe a particular kind of developmental process: induction. But the lessons from the hierarchical approach can be extended to the analysis of gene regulation in general, where these two dimensions of manipulation are present.

One dimension concerns evolutionary change, and the specificity of DNA for the conditional expression of genes. We know that mutations to regulatory regions change the context in which genes are expressed and the relationships between the expression of different genes. This is plausibly a causally specific relationship, given the wide variety of different functional relationships that can be generated by changes to the regulatory sequences.

The other dimension concerns the actual causes that are present during development, the context which drives the expression of genes via the regulation. Given the analysis of permissive causes above, what is plausibly distinctive about this causal relationship is the precision with which a particular developmental context (a cause) generates some subsequent pattern of gene expression (an effect).

If we carry this general lesson from the Waddington box back to the biology, then, it marks a change in the application of causal specificity to genes and DNA more broadly. Recent work has assumed that DNA is a specific difference-maker for particular developmental states—most notably protein synthesis. In contrast, the hierarchical approach that I used to articulate the permissive–instructive distinction suggests that we think of DNA as a specific difference-maker for the conditional

expression of proteins. The shift in focus is from genes coding for proteins to genes regulating developmental pathways.

Given the influence of Woodward's (2010) paper, it is surprising that philosophical discussion on the specificity of genes has been limited to ideas about coding sequences. Woodward's only reference to a biologist concerning the specificity of genes was Eric Davidson, whose work has focused almost solely on gene regulation and its evolution (Hinman 2016). Davidson, along with his co-author Douglas Erwin, even took part in a highly prominent debate in the journal, *Science*, where they argued for the importance of changes in regulatory sequences in contrast to coding sequences for explaining the evolution of animal body plans (Davidson and Erwin 2006; Erwin and Davidson 2006; Coyne 2006). It is plausible, then, that the specificity of interest to Davidson is the specificity of regulatory regions for the conditional expression of genes, rather than the specificity of coding regions for protein synthesis.

Here is a passage from Davidson's last book, co-written with Isabelle Peters, which conveys the kind of specificity he has in mind:¹⁰

Most of the functions of the basal transcription apparatus, which consists of an enormous complex of over 50 polypeptides, are not specific to particular genes but instead are generally deployed for gene transcription. The answers to the question why any particular gene is expressed in a specific developmental context therefore usually do not lie within the structure of the basal transcription apparatus. Rather, developmental specificity resides in the encoded combination of transcription factor target sites that constitute the working sequences of CRMs (Peter and Davidson 2015, p. 49).

A "CRM" here is a cis-regulatory module—a sequence of regulatory DNA which binds certain transcription factors and controls (along with other CRMs) when and where a gene will be turned on in a developing embryo. Note that within this short passage the term specificity occurs once, and the term specific occurs twice. What should we make of this passage in light of the analysis of the permissive–instructive distinction I have given above? There is a danger in such interpretive work, but since such passages are what lead philosophers to be interested in these different properties of causal relationships, attempting to interpret them seems appropriate here.

Although Davidson and Peters are not talking about permissive and instructive causes, there is still a clear division between two dimensions of manipulation. It will help to separate these. What is being manipulated over the developmental time scale is what Davidson and Peter refer to as the *developmental context*. It is these different contexts that actually vary over developmental time and cause a particular gene to be expressed or not. When the question of why the gene is expressed in that context arises, however, Davidson refers to the specificity "residing in" the regulatory regions in the DNA—these are the "target sites" on the DNA. Clearly, the actual DNA remains unchanged over the developmental time scale, so for it to do any explanatory work, Davidson and Peter must have in mind some counterfactual

¹⁰ The quote is similar to that cited by Woodward (2010, 302)

difference-making capacity here, or the capacity for manipulation over the evolutionary time scale. This shift from a developmental to evolutionary context to explain some particular effect mirrors what we saw in the Waddington box. Recall that the slots have a distinctive causal relationship with the buckets, but to explain this distinctive relationship we advert to the causal specificity of the pin layouts.

At this point, we can benefit from the distinction between specificity and precision given in the previous section. I suggested that what was distinctive about the causal relationship between the slots and the buckets was that it was precise, rather than specific, and that this particular kind of causal relationship may have been subsumed under the notion of specificity. Notice that there is a linguistic ambiguity here too: the word “specific” in English can also mean “precise”. One plausible interpretation here, given this insight, is that when Davidson and Peters talk of a gene being expressed in a “specific developmental context”, what they are referring to is the precision of the regulatory apparatus, rather than it being a fine-grained cause. Consider, for example, their use of specific in another passage:

This [genomic] program includes the information to read the regulatory situation at each point in time everywhere in the organism throughout development, and to utilize these newly assessed data to generate the specific new situations needed for developmental progress in exactly the correct manner (Peter and Davidson 2015, p. 3).

Here, the subject is the actual causes of gene expression in development. The word “precise” could easily replace “specific”, given that the effect is both exact and correct. If this is correct, then Peter and Davidson are saying that the *causal specificity* of transcription factor target sites is responsible for the *causal precision* of the resulting gene expression. This reflects the hierarchical analysis given in this paper: where a highly specific cause explains why another cause is precise.

In addition, Peter and Davidson contrast the specificity of these regulatory regions with that of the “basal transcription apparatus”. The claim that this apparatus is general, in contrast to the varied “target sites” suggests that we should view this particular comparison as a competitive causal analysis: the regulatory regions explain why particular genes are expressed because they are causally specific (measured over an evolutionary time-scale), whereas the basal transcription apparatus is not. Here, then, the argument resembles that made by Weber (2013, 2016) except that the DNA is now specific for the conditional expression of genes and in particular, allowing for them to be tuned to precisely control developmental pathways.

In summary, the notion of a hierarchical analysis, the introduction of two different dimensions of manipulation, and the distinction between specificity and precision, all provide insights that can help unravel a set of causal claims made about the complex interaction between development and evolution.

Conclusion

The aim of the paper was twofold. First, to attempt to clarify the elusive distinction between permissive and instructive causes by applying recent work on different properties of causal relationships in philosophy. I take it my account captures some central elements of the distinction, even if it turns out not to capture everything that biologists have ever meant when using the distinction. Second, I wanted to go beyond what I called a ‘competitive’ approach to analysing the interaction of causes. I focused on specificity, which captures the amount of *control* a causal variable has over some effect. But I argued that this effect may be another causal relationship. Exploring this hierarchical relationship between causes provides a novel way to think about the role that specificity plays in biology—applying it to the conditional expression of genes, rather than the specificity of genes for their products.

There is, however, another lesson here—a more subtle one that arises when we use the Waddington box to direct our thinking. Other recent work on causal specificity has supposed that the payoff of having fine-grained control is ultimately to be able control the production of a wide variety of effects. A focal example in the debate has been contrasting the specificity of coding sequences in DNA with post-transcriptional splicing factors (Griffiths et al. 2015; Weber 2013, 2016). The possible range of different coding sequences and the possible combinations of splicing factors multiply together to produce a huge range of possible proteins, and the production of variety is the assumed advantage of this use of highly specific control.

But the Waddington box (and by implication, the causal relationship between permissive and instructive cause) did not behave in this way—a key role that the specificity of the pin layouts played was to reduce the number of available buckets. In this case, the advantage of having fine-grained control is the opposite of the above—it is to reduce unwanted variety. This variety might simply be noise (as it was in the Waddington box). But an alternative way to interpret this decrease as the stabilisation of the effects of a cause across different environmental backgrounds or different internal contexts. Thus, the tuning of an instructive cause may make a particular inductive reaction sensitive to some inputs, but robust against other concurrent processes occurring around it. Such ideas are not new—but the idea that causal specificity may play a role in explaining them is.

Lastly, a word about DNA being special. The current emphasis on the specificity of the coding regions of DNA surely gets something right (though I suspect more work needs to be done to make this clear). But even if a bulletproof case can be made for genes being the most specific causes of proteins, this is a long way from saying they have causal specificity for anything beyond that [see Godfrey-Smith (2000) for a discussion of a similar limitation on the notion of genes ‘coding’ for outcomes]. The shift to thinking of DNA as being specific for regulation, however, allows us to push the idea that genes have high causal specificity further, to apply it to the complex process of development, rather than merely to protein synthesis. Notice, however, that although this may underwrite the idea that DNA is specific for

controlling developmental pathways or developmental mechanisms, it still falls short of stating that DNA is specific for the phenotype.

Acknowledgements I am indebted to the Theory and Method in Biosciences group at the University of Sydney, in particular to Stefan Gawronski, Paul Griffiths, Arnaud Pocheville, and Karola Stotz, and for their feedback and assistance. Funding was provided by Swansea University Templeton World Charity Foundation.

Appendix

The appendices give further details for constructing the Waddington box and generating the distributions that were analysed in the paper. The code for producing this (and other) Waddington boxes, and the definitive source for how the simulations work, is freely available at <https://github.com/brettc/waddington-box>.

The layouts

There are three aims in limiting the way pins can appear in a layout. The first is to reduce the total number of layouts to something that can be analysed in a reasonable length of time. The second aim is ensure that all layouts permit the ball to reach the buckets (in combination with the physics rules defined below). The third aim is to ensure the layouts produce something interesting—a variety of different mappings.

I consider the layout of individual rows of pins first, as the full number of layouts will be constructed by combining these. The top row is a special case, which I will discuss last. The remaining four rows each contain 15 holes which could contain pins. The unrestricted number of possible pins layouts for such a row is thus $2^{15} = 32,768$. I now add two restrictions.

- (L1) Pins must have at least three holes between them, and
- (L2) there must be at least three pins in each row.

This reduces the number of possible pin configuration in a single row to 99.

Now for the top row. Why is it different? I could have just had five rows with 15 holes, but the placement of the slots means that many of the pins on the top row will never interact with the ball. Instead, the top row has two sets of three holes below the slots, and each set can contain zero or one pins. So the total layouts for the top row are $4^2 = 16$. The total number of layouts with these five rows is $16 \times 99^4 = 1,536,953,616$. To further reduce this, I add one last restriction:

- (L3) A pin can never lie directly above another pin on the next row.

This brings the total number of layouts down to 3,306,664, which is small enough for an exhaustive analysis.

The physics

To implement a simple physics that captures a ball falling through a series of rows with pins I make the following assumptions:

- (R1) A ball falls in one of 15 discrete channels that line up with the 15 pins.
- (R2) A ball interacts with each row independently, arriving at the row in one channel and exiting at one or two channels.

The way that a ball interacts with the pins on each row obeys the following rules:

- (R3) *The ball glances off a pin* If there is a pin in the channel to the right of the ball it shifts one channel left; if there is a pin to the left it shifts one channel to the right.
- (R4) *The ball bounces on a pin; it could go either way* If there is a pin in the same channel as the ball it takes two paths, moving left two channels with probability 0.5 and right two channels with probability 0.5.
- (R5) *The ball hits the edge of the box* If either of the previous two rules would cause the ball to leave the 15 channels, the ball goes to two channels to the opposite of the pin it encountered.

Rule (R4) means that the path a ball takes can split, and we must then track both balls as they descend. A path can split multiple times and carries with it a probabilities. These probabilities are then accumulated in the buckets. The four buckets, $\{b_1 \dots b_4\}$ collect the ball falling in the following channels:

- (R6) *The four buckets collect from the 15 channels* $b_1 \leftarrow (1, 2, 3, 4)$,
 $b_2 \leftarrow (4, 5, 6, 7, 8)$, $b_3 \leftarrow (8, 9, 10, 11, 12)$, $b_4 \leftarrow (12, 13, 14, 15)$

Notice that the buckets overlap—some channels end up in both buckets. This works the same as if they bounced on a pin: with a probability of 0.5 that they go either way.

The splitting of pathways is the key to Galton's box. A ball falling on top of a pin has an equal probability of going left or right. If the ball falls directly onto another pin, the pathways split again. In the Quincunx layout of Galton's box, the pathways produce a binomial distribution across the buckets at the bottom (which approximates a normal distribution). In the Waddington box, this is what allows the layouts to generate a range of different probability distributions over the buckets.

References

- Bhattacharya S, Zhang Q, Andersen ME (2011) A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst Biol* 5:85. doi:[10.1186/1752-0509-5-85](https://doi.org/10.1186/1752-0509-5-85)
- Bulmer MG (2003) Francis Galton: pioneer of heredity and biometry. JHU Press, Baltimore

- Coyne JA (2006) Comment on gene regulatory networks and the evolution of animal body plans. *Science* 313(5788):761. doi:[10.1126/science.1126454](https://doi.org/10.1126/science.1126454) (author reply 761)
- Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. *Science* 311(5762):796–800. doi:[10.1126/science.1113832](https://doi.org/10.1126/science.1113832)
- De Robertis EM (2006) Spemann’s organizer and self-regulation in amphibian embryos. *Nat Rev Mol Cell Biol* 7(4):296–302. doi:[10.1038/nrm1855](https://doi.org/10.1038/nrm1855)
- Erwin DH, Davidson EH (2006) Response to comment on “gene regulatory networks and the evolution of animal body plans”. *Science* 313(5788):761c–761c. doi:[10.1126/science.1126765](https://doi.org/10.1126/science.1126765)
- Gilbert SF (1991) Epigenetic landscaping: Waddington’s use of cell fate bifurcation diagrams. *Biol Philos* 6(2):135–154
- Godfrey-Smith P (2000) On the theoretical role of “genetic coding”. *Philos Sci* 67(1):26–44. doi:[10.1086/392760](https://doi.org/10.1086/392760)
- Griffiths PE, Gray RD (2005) Discussion: three ways to misunderstand developmental systems theory. *Biol Philos* 20(2–3):417–425. doi:[10.1007/s10539-004-0758-1](https://doi.org/10.1007/s10539-004-0758-1)
- Griffiths PE, Pocheville A, Calcott B, Stotz K, Kim H, Knight R (2015) Measuring causal specificity. *Philos Sci* 82(4):529–555. doi:[10.1086/682914](https://doi.org/10.1086/682914)
- Hinman VF (2016) Eric Davidson (1937–2015) and the past, present and future of EvoDevo. *Evol Dev*. doi:[10.1111/ede.12180](https://doi.org/10.1111/ede.12180)
- Holtzer H (1968) Induction of chondrogenesis. A concept in terms of mechanisms. In: Glieschmajer R, Billingham R (eds) *Epithelial-mesenchymal interactions*. William and Wilkins, Baltimore
- Huang S (2012) The molecular and mathematical basis of Waddington’s epigenetic landscape: a framework for post-Darwinian biology? *BioEssays* 34(2):149–157. doi:[10.1002/bies.201100031](https://doi.org/10.1002/bies.201100031)
- Kirschner MW, Gerhart JC (2006) *The plausibility of life: resolving Darwin’s dilemma*. Yale University Press, New Haven
- Oyama S (2003) Terms in tension: what do you do when all the good words are taken? In: Oyama S, Gray RD, Griffiths PE (eds) *Cycles of contingency: developmental systems and evolution*, reprint edn. A Bradford Book, Cambridge
- Oyama S, Gray RD, Griffiths PE (eds) (2003) *Cycles of contingency: developmental systems and evolution*, reprint edn. A Bradford Book, Cambridge
- Peter I, Davidson EH (2015) *Genomic control process: development and evolution*, 1st edn. Academic Press, Boston
- Pocheville A, Griffiths PE, Stotz K (2017) Comparing causes: an information-theoretic approach to specificity, proportionality and stability. In: *Proceedings of the 15th congress of logic, methodology and philosophy of science*. College, London
- Siegal ML, Bergman A (2002) Waddington’s canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci USA* 99(16):10,528–10,532. doi:[10.1073/pnas.102303999](https://doi.org/10.1073/pnas.102303999)
- Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150(3811):563–565
- Wang J, Xu L, Wang E, Huang S (2010) The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys J* 99(1):29–39. doi:[10.1016/j.bpj.2010.03.058](https://doi.org/10.1016/j.bpj.2010.03.058)
- Wang J, Zhang K, Xu L, Wang E (2011) Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc Natl Acad Sci* 108(20):8257–8262. doi:[10.1073/pnas.1017017108](https://doi.org/10.1073/pnas.1017017108)
- Waters CK (2007) Causes that make a difference. *J Philos* 104:551–579
- Weber M (2013) Causal selection versus causal parity in biology: relevant counterfactuals and biologically normal interventions. In: *What if? On the meaning, relevance and epistemology of counterfactual claims and thought experiments*. University of Konstanz, Konstanz, pp 1–44
- Weber M (2016) Discussion note: which kind of causal specificity matters biologically. *Philos Sci*. doi:[10.1086/692148](https://doi.org/10.1086/692148)
- Wilkins AS (2002) *The evolution of developmental pathways*. Sinauer Associates, Sunderland
- Wilkins AS (2008) Waddington’s unfinished critique of neo-Darwinian genetics: then and now. *Biol Theory* 3(3):224–232
- Woodward J (2010) Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biol Philos* 25(3):287–318. doi:[10.1007/s10539-010-9200-z](https://doi.org/10.1007/s10539-010-9200-z)