Social Learning and the Baldwin Effect
David Papineau

1 Introduction

The Baldwin effect occurs, if it ever does, when a biological trait becomes innate as a result of first being learned. Suppose that some trait is initially absent from a population of organisms. Then a number of organisms succeed in learning the trait. There will be a Baldwin effect if this period of learning leads to the trait becoming innate throughout the population.

Put like that, it sounds like Lamarckism. But that is not the idea. When James Mark Baldwin and others first posited the Baldwin effect over a hundred years ago, their concern was precisely to uncover a respectable Darwinian mechanism for the Baldwin effect [1[1]]. The great German cytologist Augustus Weismann had already persuaded them that there is no automatic genetic inheritance of acquired characteristics: the ontogenetic acquisition of a phenotypic trait cannot in itself alter the genetic material of the lineage that has acquired it. The thought behind the Baldwin effect is in effect that an alternative Darwinian mechanism might nevertheless mimic Lamarckism, in allowing learning to influence genetic evolution, but without requiring Lamarck's own discredited hypothesis that learning directly affects the genome.

Why should we be interested in the possibility of Baldwin effects? One reason the topic attracts attention is no doubt that it seems to soften the blind randomness of natural selection, by allowing the creative powers of mind to make a difference. Still, there are other good reasons for being interested in the Baldwin effect, apart from wanting some higher power to direct the course of evolution.

Consider the many innate behavioural traits whose complexity makes it difficult to see how they can be accounted for by normal natural selection. I have in mind here innate traits that depend on a number of components that are of no obvious advantage on their own. For example, woodpecker finches in the Galapagos Islands use twigs or cactus spines to probe for grubs in tree braches. This behaviour is largely innate (Tebbich et al 2001). It also involves a number of different behavioural dispositions—finding possible tools, fashioning them if necessary, grasping them in the beak, using them to probe at appropriate sites—none of which would be any use by itself. For example, there is no advantage in grasping tools if you aren't disposed to probe with them, and no advantage to being disposed to probe with tools if you never grasp them. Now, insofar as the overall behaviour is innate, these different behavioural components will presumably depend on various independently inheritable genes. However, this then makes it very hard to see how the overall behaviour can possibly be selected for. In order for the behaviour to be advantageous, all the components have to be in place. But this will require that all the relevant genes be

[1[1]] In the 1890s Henry Osborne and Conwy Lloyd Morgan had also proposed that non-Lamarkian processes could lead to acquired characteristics becoming innate. Given this, 'Baldwin has done well to have become the namesake for the effect', as Peter Godfrey-Smith (2003) observes. The familiar term 'The Badwin Effect' is due to George Gaylord Simpson's 1953 article of that title, which ironically was largely concerned to belittle the effect. For much more interesting history of the Baldwin effect, see Griffiths (2003).

present together. However, if these are initially rare, it would seem astronomically unlikely that they would ever co-occur in one individual. And, even if they did, they would quickly be split up by sexual reproduction. So the relevant genes, taken singly, would seem to have no selective advantage which would enable them to be favoured by natural selection.

But now add in the Baldwin effect. This now promises a way to overcome the selective barrier. We need only suppose that some individuals are occasionally able to acquire the behaviour using some kind of general learning mechanism. If they can succeed in this, then the Baldwin effect can kick in, and explain how the behaviour becomes innate. Thus behaviours whose selection seems mysterious from the point of view of orthodox natural selection can become explicable with the help of the Baldwin effect.

But I am getting ahead of myself. This last suggestion assumes that the Baldwin effect is real, and that has yet to be shown. In the rest of this paper I shall explore possible mechanisms for the Baldwin effect, and consider whether they may be of any biological significance. My general verdict will be positive. I shall aim to show how there are indeed mechanisms which can give rise to Baldwin effects, and moreover that there is some reason to think that such Effects have mattered to the course of evolution.

I became interested in the Baldwin effect because it has always seemed to me obvious that there is at least one kind of case where it operates—namely, with the social learning of complex behavioural traits. It will be helpful to consider this in broad outline before we get caught up in analytic details. Suppose some complex behavioural trait P is socially learnt—individuals learn P from others, where they have no real chance of figuring it out for themselves. This will then create selection pressures for genes that make individuals better at socially acquiring P. But these genes wouldn't have any selective advantage without the prior culture of P, since that culture is in practice necessary for any individual to learn P. After all, there won't be any advantage to a gene that makes you better at learning P from others, if there aren't any others to learn P from. So this then looks like a Baldwin effect: genes for P are selected precisely because P was previously acquired via social learning.

By way of an example, consider the woodpecker finches again, and suppose that there was a time when their tool-using behaviour was not innate but socially learned[2[2]]. That is, young woodpecker finches would learn how to use tools from their parents and other adepts. Now, this socially transmitted culture of tool use would give a selective advantage to genes that made young finches better at learning the trick. For example, it would have created pressure for a gene that disposed finches to grab suitable tools if they saw them, since this would give them a head start in learning the rest of the grub-catching behaviour from their elders. But this gene wouldn't have been advantageous on its own, in the absence of the tool-using culture, since even finches with that gene wouldn't have been able to learn the rest of the tool-using behaviour, without anyone to teach them.

[2[2]] There are well-evidenced examples of tool use being transmitted culturally in other birds and in primates. Hunt and Gray (2003), Whiten et al. (1999).

In what follows I shall be particularly interested in cases of this kind—that is, cases where social learning gives rise to Baldwin effects. From the beginning, theorists have often mentioned social learning in connection with the Baldwin effect, but without pausing to analyse its special significance. (For an early example, see Baldwin himself, 1896; for a recent one, see Watkins, 1999.) I shall offer a detailed explanation of the connection between social learning and Baldwin effects. As we shall see, there are two main biological mechanisms that can give rise to Baldwin effects—namely, 'genetic assimilation' and 'niche construction'. Social learning has a special connection with the Baldwin effect because it is prone to trigger both of these mechanisms. When we have social learning, then we are likely to find cases where niche construction and genetic assimilation push in the same direction, and thus create powerful biological pressures.

Much recent literature argues that, while there are indeed biological processes that fit the specifications of the Baldwin effect, it is a mistake to highlight the Baldwin effect itself as some theoretically significant biologically mechanism. (Cf. Downes, 2003, Griffiths, 2003.) Rather, Baldwin-type examples are simply special cases of more general biological processes. In particular, they are special cases of either genetic assimilation or niche construction. This is a perfectly reasonable point. As we shall see, genetic assimilation and niche construction are the two main sources of Baldwin effects, and both of these processes are of more general significance, in that they don't only operate in cases where a learned behaviour comes to be innate, but in a wider range of cases, many of which may involve neither learning nor behaviour.

Still, if we focus on the social learning cases I am interested in, then the Baldwin effect re-emerges as a theoretically important category. These cases are important, as I said, precisely because they combine both niche construction and genetic assimilation. This combination gives rise to particularly powerful biological pressures, and for this reason is worth highlighting theoretically. Moreover, this combination of pressures arises specifically when a socially learned behaviour leads to its own innateness, and is not found more generally. So the Baldwin effect turns out to be theoretically significant after all.

2 Preliminaries: Genetic Control, Innateness, and Social Learning

Before proceeding to analysis of the Baldwin effect itself, it will be helpful to clarify various preliminary issues. In this section I shall first discuss the selective advantages and disadvantages of having behavioural traits controlled by genes rather than learning, and then explain what I mean by 'innate' and 'social learning' respectively in the context of the Baldwin effect.

2.1 Genetic Control versus Learning

In the woodpecker finch example above, I took it for granted that it would be selectively advantageous for the relevant behaviour to depend on genes rather than learning. Since this assumption is generally required for the Baldwin effect, and since it is by no means always guaranteed to be true, it will be useful briefly to discuss the conditions under which it will be satisfied.

It might seem unlikely that there will ever be any selective advantageous to bringing some trait P under genetic control, given that it can be learned anyway. If some adaptive P is going to be acquired by learning in any case, what extra advantage derives from its genetic determination?

Well, one response is that P won't always be acquired in any case, if it is not genetically fixed. Learning is hostage to the quirks of individual history, and a given individual may fail to experience the environments required to instil some learned trait. Moreover, even if the relevant environments are reliably available, the business of learning P may itself involve immediate biological costs, diverting resources from other activities, and delaying the time at which P becomes available.

These obvious advantages to genetic fixity—reliability and cheapness of acquisition—can exert a greater or lesser selective pressure, depending on how far genetic fixity outscores learning in these respects. On the other side, however, must be placed the loss of flexibility that genetic fixity may entail. Learning will normally be adaptive across a range of environments, in each case producing a phenotype that is advantageous in the current environment. Thus, if the environment were to vary in such a way as to make P maladaptive, an organism with genes that fix P may well be less fit than one which relies on learning, since the latter would not be stuck with P, and may instead be able to acquire some alternative phenotype adapted to the new environment.

As a general rule, then, we can expect that genetic fixity will be favoured when there is long-term environmental stability, and that learning will be selected for when there are variable environments. Given environmental stability, genetic fixity will have the aforementioned advantages of reliable and cheap acquisition. But these advantages can easily be outweighed by loss of flexibility when there is significant environmental instability.

In thinking about these issues, it is helpful to think of the relevant behaviours as initially open to shaping by some repertoire of relatively general learning mechanisms (perhaps including classical and instrumental conditioning, plus various modes of social learning). The question is then whether the behavioural trait in question should be switched, so to speak, from the control of those general learning mechanisms to direct genetic control. However, perhaps it should not be taken for granted that the general learning repertoire will itself be unaffected by such switching. Maybe bringing one behavioural trait under genetic control will make an organism less efficient at learning other behavioural traits. (Cf. Godfrey-Smith, 2003.) For example, the woodpecker finches may be less able to learn to learn other ways of feeding, once their tool-using behaviour becomes genetically rigid. If so, this too will need to be factored in when assessing the selective gains and losses of bringing some behaviour under genetic control.

Exactly how the pluses and minuses of genetic control versus learning work out will depend on the parameters of particular cases.[3][3] Still, I hope it is clear enough that there will be some cases where genetic fixity will have the overall biological

---

[3][3] For a detailed quantitative analysis of the relative costs of learning and genetic control, see Mayley (1996).

advantage, even if there are other cases where learning will be biologically preferable.[4][4] So from now on I shall assume we are dealing with examples where the selective advantages of genetic control does outweigh the costs, since it is specifically these cases that create the possibility of Baldwin effects

2.2 'Innate'

So far I have been proceeding as if there were a clear distinction between 'innate' and 'acquired' traits. However, I do not think that this distinction is at all clear-cut. No definite meaning attaches to the notion of an 'innate trait', once we move away from the genome itself to any kind of phenotypic trait, since nothing outside the genome is determined by the genes alone (even the appearance of basic organs can be disrupted by non-standard environments). True, there are a number of other criteria which are widely taken to constitute 'innateness', such as presence at birth, universality through the species, being a product of natural selection, and high developmental insensitivity to environmental variation. However, these criteria all dissociate in both directions in real-life cases. Because of this, the notion of innateness can be a source of great confusion. If you ask me, far more harm than good results from unthinking deployment of this notion. (Cf. Griffiths, 2002.)

Even so, it will be convenient for the purposes of this paper to continue to talk about traits that are at one time 'acquired' later becoming 'innate'. When I use this terminology, I should be understood in terms of the last criterion mentioned above, that is, high developmental insensitivity to environmental variation. I shall take a trait to be innate to the extent that it has a 'flat norm of reaction', that is, to the extent that it reliably occurs across a wide range of developmental contexts. Note that it follows from this criterion that a trait will not be innate to the extent it is 'learned', given that learning can be understood as a mapping from different developmental environments to different phenotypes.[5][5]

Given this understanding of innateness, then, innateness comes out as a matter of degree: as observed above, no non-genomic traits have a completely flat norm of reaction, in the sense of developing in all environments; at most, we will find that some traits are less sensitive to environmental variation than others. This does not worry me. A comparative notion of innateness will be perfectly adequate for the purposes of this paper. It will be interesting enough if we find Baldwin effects where the prior learning of certain traits leads to the selection of new genes that make traits less sensitive to environmental variation, rather than absolutely insensitive. Talk about traits becoming innate should be understood in this comparative way from now on.

In this connection, it may be helpful to think of behavioural traits in terms of neural pathways in the brain. The trait will be present when appropriate sensory inputs

---

[4][4] In contexts where learning has the biological advantage over genetic fixity, then we might well find 'reverse Baldwin effects', with some trait originally under genetic control coming to depend on learning instead.

[5][5] Some favour the far more controversial thesis that not being learned is sufficient as well as necessary for innateness, at least in the context of psychological traits: that is, not only are psychological traits not innate if they are learned, but also that they are innate if they are not learned. Cf. Samuels (2002); see also Cowie (1999).

trigger relevant motor outputs. Some genomes may leave a large 'gap' between sensory and motor pathways, in which case general learning mechanisms will have plenty of work to do in closing them. Other genomes may only leave a small such gap, one that can be closed with a minimum of environmental input. However, general evolutionary considerations suggest that it will be unusual to find no gap at all. (Why bother with genes that close the gap entirely, once it is so small that nearly all normal environments will bridge it? In this connection, note that even the highly innate tool use of the Galapagos woodpecker finches still require a modicum of individual trail-and-error learning during a short critical period. (Tebbich et al., 2001.))

2.3 'Social Learning'

I shall use the term 'social learning' to cover all processes by which the display of some behaviour by one member of a species increases the probability that other members will perform that behaviour. This covers a numbers of different mechanisms, but I intend my analysis of social learning and Baldwin effects to apply to them all.

Thus we can distinguish (cf. Shettleworth, 1998, Tomasello, 2000):

(i) Stimulus Enhancement. Here one animal's doing P merely increases the likelihood that other animals' behaviour will become conditioned to relevant stimuli via individual learning. For example, animals follow each other around—novices will thus be led by adepts to sites where certain behaviours are possible (pecking into milk bottles, say, or washing sand off potatoes) and so be more likely to acquire those behaviours by individual trial-and-error.

(ii) Goal Emulation. Here animals will learn from others that certain resources are available, and then use their own devices to achieve them. Thus they might learn from others that there are ants under stones, or berries in certain trees.

(iii) Blind Mimicry. Here animals copy the movements displayed by others, but without appreciating to what end these movements are a means. While it is possible that some non-human animals can do this, it seems to be a relatively high-level ability.

(iv) Learning about Means to Ends. Here animals grasp that some conspecific's behaviour is a means to some end, and copy it when they want that end. There is little evidence that non-human animals can do this (but see Akins and Zentall, 1998).

These processes differ in significant ways. For example, (i) and (ii), unlike (iii) and (iv), do not lend themselves to cumulative culture, since any technical sophistication developed by one individual will not be passed on to the others who duplicate their behaviour (Boyd and Richerson, 1996, Tomasello, 2000). Again, (iii)—blind mimicry—but not the other modes of social learning, is highly sensitive to which individuals are taken as models, since in this case there is no further mechanism to ensure that only adaptive behaviours are copied. Differences like these may well interact interestingly with the Baldwin effect. However, I shall not pursue these complexities in what follows (though see footnote 10 below). I shall simply assume the general definition of social learning given above, and stick to points that apply to all its species.

3 Why Does Learning Matter?

In an extremely illuminating article on the Baldwin effect, Peter Godfrey-Smith schematises the structure of the Effect roughly as follows (Godfrey-Smith, 2003).

Stage 0 The environment changes so as to make phenotype P adaptive.

Stage 1 Some organisms learn P and prosper accordingly.

Stage 2 There is selection of genes which make P innate.

Given this schematisation [6][6], Godfrey-Smith then raises the obvious question about the Baldwin effect. Why should going through Stage 1 be crucial to reaching Stage 2? Why do we need any learning of P en route to the selection of genes for P? After all, Stage 0 already ensures that organisms with P will survive better, and thus on its own would seem to guarantee that genes for P will have a selective advantage, whether or not there is any intermediate learning. So won't Stage 2—selection of genes for P—be triggered immediately by Stage 0—phenotype P becomes adaptive, without any necessity of a detour through Stage 1?

This worry is widely taken to show the Baldwin effect is a chimera. John Watkins (1999), for example, has recently argued that the Baldwin effect is impossible on precisely these grounds. Watkins allows that any organisms that do acquire P by learning will on that account be more likely to survive and pass on their genes. However, he points out, there is no reason to suppose that those organisms are especially likely to have the genes that make P innate, and so their surviving by learning P would seem to contribute nothing to the selection of those genes.

However, Watkins is too quick to dismiss the possibility of Baldwin effects. Despite his argument, there are various special cases where the selection of genes for P may indeed depend on P previously being learned. Following Godfrey-Smith, I shall consider three possible such cases, which I shall call 'Breathing Spaces', 'Niche Construction', and 'Genetic Assimilation' respectively.

The first suggestion—breathing spaces—is simply the idea that populations of organisms may not survive long enough to allow the selection of the genes for P, if they are not able to learn P in the interim. This seems to have been Baldwin's own thought.[7][7] Some environmental changes may be so drastic that the populations which undergo them will face extinction if they cannot adapt quickly. In the face of such drastic environmental changes, learning may allow a significant number of organisms to acquire the necessary adaptive trait P, at a time where genes determining

---

[6][6] Godfrey-Smith also requires selection of genes for learning at Stage 1. This seems to me an unhelpful restriction, fostered by an excessive focus on genetic assimilation (sections 4 and 5 below). Pace Godfrey-Smith, there need be no Stage 1 selection of genes for learning in niche construction cases of Baldwin effects (sections 6 and 7).

[7][7] See Baldwin (1896), section II. In addition, there are indications (section III.2) that Baldwin also thought that the learned predominance of a trait would lead to sexual preferences for displays of that trait; this would be a case of niche construction rather than breathing spaces. Cf. Griffiths (2003) section 3.

P are still rare. This would then allow the population to stay around long enough for natural selection to drive the genes for P to fixity.

It is doubtful whether breathing spaces are of any real biological significance. Few environmental changes seem likely to fit its requirements. There are certainly plenty of environmental changes that destroy whole populations—the impact of an asteroid, the commercial destruction of a rain forest—but these are not the kind of catastrophes that can be averted by learning new adaptive tricks. Conversely, environmental changes that are gradual enough to allow organisms to learn new tricks—climatic shifts, say, or the immigration of a new predator—will rarely be so urgent that the whole population would be under threat of complete extinction without the tricks, in which case there will be time for genetic selection to operate even without a learning stage. Considerations such as these lead Godfrey-Smith to dismiss breathing spaces as of dubious importance, and I agree with him. I shall say no more about breathing spaces in this paper.

That leaves niche construction and genetic assimilation. With niche construction, the idea is that Stage 1 alters selective pressures so as render genes for P advantageous, when they weren't in Stage 0. With genetic assimilation, by contrast, the learning of P doesn't alter selection pressures; rather it is itself the function that renders certain genes advantageous. Both these processes require extended discussion. It will be convenient to begin with genetic assimilation, which will occupy the next two sections. After that I shall return to niche construction.

4 The Baldwin Effect as Genetic Assimilation

The notion of genetic assimilation is due to C.H. Waddington. In the 1940s and 1950s he investigated the way in which the selection of traits triggered by special environments could lead to those traits developing automatically across a wide range of environments. Waddington applied this idea to biological development in general, not just to behavioural traits that are initially acquired by learning. Still, our immediate concern here is with possible mechanisms for Baldwin effects, and so I shall focus on this kind of case, returning to Waddington's wider concerns in the next section.

Let me introduce the logic of genetic assimilation by considering a simple model. Suppose n sub-traits, P_i, i = 1, . . ., n, are individually necessary and jointly sufficient for some adaptive behavioural phenotype P. (You need to be able to find tool materials, fashion them, grasp them, . . . As before, each individual sub-trait is no good without all the others.) Each sub-trait can either be genetically fixed or acquired through learning. (For this section's purposes there is no need to assume that this will be social learning—any mode of learning will do.) Suppose further that each sub-trait is under the control of a particular genetic locus: one allele at this locus will genetically determine the sub-trait, while an alternative allele leaves the sub-trait plastic and so available for learning. So, for sub-trait P_i, we have allele I_G which genetically fixes P_i, and allele I_L which allows it to be learned.

To start with, the I_Gs that genetically determine the various P_is are rare, so that it is highly unlikely that any individual will have all n P_is genetically fixed. Moreover, suppose that it is pretty difficult to get all n P_is from learning. Still, given these

specifications, organisms that have some P_is genetically fixed will face less of a task in learning the rest. (If you are already genetically disposed to grab suitable twigs if you see them, you will have less to do to learn the rest of the tool-using behaviour.) Organisms who already have some I_Gs will have a head start in the learning race, so to speak, and so will be more likely to acquire the overall phenotype. So the I_Gs that give them the head start will have a selective advantage over the I_Ls. Natural selection will thus favour the I_Gs over the I_Ls, and in due course will drive the I_Gs to fixity. The population will thus move through a stage where P is acquired by learning (Stage 1) to a stage where it is genetically fixed (Stage 2), thus yielding a prima facie Baldwin effect.

This model is a simplification of one developed by Hinton and Nowlan (1987). They ran a computer simulation of essentially the above structure using a 'sexually reproducing' population of neural nets, and showed that the dynamics of their simulation would indeed progressively replace the alleles I_L which left the P_is to learning with the I_Gs that fixed them genetically.

To better see what is going on in this model, consider the standard worry about the natural selection of a complex of genes none of which is any good on its own. Thus: 'What is the advantage of any I_G on its own, given that it only fixes one P_i, which isn't of any use without the other n-1 P_is? Don't we need all n I_Gs to occur together for any of them to yield a biological advantage? But that is overwhelmingly unlikely, if they are initially rare, and anyway they would be split up by sexual reproduction, if they did ever co-occur. So each I_G on it own would seem to have no selective advantage.'

The above model allows a cogent answer to this argument: each I_G does have a selective advantage on its own, even in the absence of the other I_Gs, precisely because it makes it easier to learn the rest of P. Even in the absence of other I_Gs at other loci, any given I_G will still be favoured by natural selection, because it will reduce the learning load and so make it more likely that its possessor will end up with the advantageous phenotype P. This is what drives the progressive selection of the I_Gs in the model. Each I_G is advantageous whether or not there are I_Gs at other loci, simply because having an I_G rather than an I_L at any given locus will reduce the amount of further learning needed to get the overall P.

Given this last point, it will be worth thinking a bit about the precise sense in which the modelled process would constitute a Baldwin effect. If we focus on any specific locus, it turns out that the prior learning allowed by that locus's I_L is unnecessary for the selection of the I_G after all. This I_G will have an advantage over its alternative I_L quite independently of any such prior learning. For the possession of this I_G by any given individual will reduce its overall learning load, by removing component P_i from the vagaries of learning and placing it under genetic control. This remains true whatever alleles are at other loci, and even if no organisms have ever previously used the alternative I_L to learn P_i. So from this perspective the Baldwin effect seems to have disappeared. Stage 1, in which the organisms learn P_i, seems to play no role in fostering the selection of I_G, just as John Watkins suspected.

In order to see why the genetic assimilation model does indeed deliver a Baldwin effect, we need to adopt a wider perspective, and consider the progressive accumulation of—not just one specific I_G —but of all the I_G alleles, at different loci,

which contribute to the overall genetic fixity of P. Recall that in the early stages of this process the various I_Gs are rare. This means that, even if a lucky individual does have one or two I_Gs, any success in acquiring the overall P will depend on its learning the remaining P_is. Moreover, it is precisely this possibility that gives the various I_Gs their initial selective advantage. Any given I_G is advantageous precisely because of the way it makes it more likely the organism will be able learn the remaining non-innate components of P.

So now we do have a story in which learning matters. The progressive selection of the whole complex of alleles hinges on the fact that organisms are able to learn elements of P. We wouldn't arrive at the final stages, where all the P_is get genetically fixed, were it not that in the early and intermediate stages the organisms were able to learn non-innate components of P—otherwise, to repeat, none of the I_Gs would have any initial selective advantage. So, if we consider the overall accumulation of I_Gs, we will observe a sequence, with each stage causally necessary for the next, in which learned components of P are progressively replaced by innate ones. The behaviour P becomes innate as a result of a process in which P's earlier being learned plays an essential role. The process is thus a Baldwin effect.

5 Waddington and Genomic Space

In this section I want to compare the model just outlined with Waddington's original notion of genetic assimilation. In the 1940s and 50s Waddington and others were interested in 'canalization', that is, in the buffering of adaptive traits against disruptive environments. To the extent that some trait is highly important to fitness—having normal hands, say—it will be advantageous that it should appear across a wide range of environments, including various non-standard ones. Because of this, Waddington and his associates wondered whether there could be selection for canalisation. Could natural selection operate in favour of genomes which 'flatten norms of reactions' of important traits—that is, which ensure that these traits will appear in a wider range of environments that hitherto?

In a famous series of experiments, Waddington demonstrated that this could indeed happen. For example, he subjected a population of fruit fly embryos to heat shocks. As a result, some failed to grow cross-veins on their wings (he called this trait 'veinless'). By breeding from these individuals, he was able to select a strain that displayed the veinless trait even in the absence of early heat shocks. (Waddington, 1953)

Waddington called his 1953 paper the 'Genetic Assimilation of an Acquired Character'. At first some of the fruit flies acquire the characteristic 'veinless' as a result an environment of heat shocks. Then, under the artificial selective regime of the experiment, in which only flies with the veinless phenotype survive, we find flies that are innately veinless and can be further selected.

Let us compare Waddington's experiment with the model of the previous section. One obvious difference is that veinless is a morphological trait, rather than a behavioural one. Moreover, it is only 'learned' in the extremely attenuated sense in which any environmentally dependent trait is 'learned'; certainly its acquisition is not the upshot of any general learning mechanism. However, let us put these relatively

inconsequential differences to one side, and focus on the question of whether Waddington's examples display the same underlying mechanism as modelled in the last section.

On the surface, they certainly do not. Veinlessness is not composed of a number of sub-traits, like some complex sequence of behaviour. Nor, correspondingly, is there anything in Waddington's analysis about a number of genetic loci, each of which can either innately determine some sub-trait, or leave it to environmental factors.

However, precisely because of these differences, Waddington's examples are puzzling in a way that the kind of case modelled in the last section is not. As has often been pointed out, there is no intrinsic reason why selecting flies that are veinless-if-heat-shocked should yield a population with an increased likelihood of innately veinless flies. Think of the flies as having three types of genome: those that ensure they have cross-veins even if heat shocked; those that make them veinless-if-heat-shocked; and those that render them innately veinless. Most of the flies in the original population had the first genome. By subjecting them to heat shocks and selecting for veinlessness we get a population with the second genome. Now, why should the third genome be more probable in the second population than in the first? Why, so to speak, should the second and third genomes' similarity in phenotypic space—they are both capable of displaying veinlessness—mean that they are similar in genomic space—a population with the second genome makes the appearance of the third more likely? (Cf. Mayley, 1996.) Why, to revert to our original terms, should Stage 1, in which the trait is 'learned', be crucial to reaching Stage 2, where it is innate?

Well, here is one possible explanation. Suppose veinlessness depends on two factors: (i) some developmentally important protein loses its required conformation, and (ii) the 'heat shock protein' needed to correct this is absent. Both of these factors can be genetically determined, but normal flies lack the requisite genes. Now think of environmental heat shocks as an alternative non-genetic way of causing the deformation in the developmental protein. Even when this happens, most flies have the heat shock protein required to remedy this. Some, however, have a genetic abnormality that means they lack the heat shock protein. These flies will be distinguished by displaying veinlessness if heat shocked. So selecting these latter flies yields a population that innately lacks the heat shock protein. Thus any flies in this new population would end up innately veinless if they were also to have a genetic abnormality that deformed the developmental protein without heat shocks. By contrast, this further abnormality would not produce innate veinlessness in normal flies, since they do have the heat shock protein which remedies the deformation.

This story now explains, in roughly the style of the last section's model, why selecting for the phenotype veinless-if-heat-shocked should make the genome innately veinless more accessible. Two factors are necessary to make a fruit fly veinless: a deformed developmental protein, and a lack of the heat shock protein. Even so, the gene that determines no heat shock protein is 'advantageous' on its own, precisely because it makes it easier to end up veinless, since it ensures veinlessness in those organisms who are subject to environmental heat shocks. And so, once this 'advantage' has led

to the selection of the gene for no heat shock protein, we only need one further gene, not two, to get innate veinlessness.[8[8]]

True, the match between this explanation and the last section's model remains inexact. For one thing, my suggested explanation does not regard veinlessness itself as composed as a number of sub-traits; rather the explanation works by factoring the determinants of veinlessness into independent components, not veinlessness itself. In addition, while the last section's model assumed innate and environmental alternatives (I_L or I_G) for all the components of P, my explanation of veinlessness only had factor (i)—deformation of the developmental protein—as environmentally acquirable; the heat shock protein was either innately present or innately absent.

This shows that the model of the last section is rather more restrictive than is necessary to capture the underlying selective dynamics. It is not essential that the trait at issue itself factors into independent sub-traits, as long it is causally depends on various independent factors that are individually necessary and jointly sufficient. Nor is it required that all these factors are open to environmental as well as genetic causation; there may be some initial genes which have a selective advantage because they mean that the trait will appear as soon as the environment supplies the other factors, even if the factors that these genes determine cannot themselves be environmentally caused.

Still, now I have discussed Waddington's own cases, it will be convenient to return to the more restrictive model of the last section, as it applies to the examples which matter to this paper. So when I talk about 'genetic assimilation' in what follows, I shall be referring to cases in which some behavioural trait itself decomposes into sub-traits, each of which can either be environmentally of genetically determined, and where the initial selective advantage of all these genes derives from the fact that it makes it easier to learn the rest of the behaviour.

6 Niche Construction

With genetic assimilation, prior learning of the trait P does not alter the selective pressures on the genes that might render P innate. Rather, enhanced learning is the function which renders those genes advantageous. At any stage in the genetic assimilation of P, these genes are preferable to alternative alleles because they make it more likely the rest of P will be learned—and this advantage does not depend on such learning previously having occurred, only on its being possible henceforth.

[8[8]8[8]] Does this second gene arise from mutations in the 'veinless-if-heat-shocked' population, or was it always present in the original experimental population? It is sometimes observed that there wouldn't have been enough time in Waddington's experiments for the relevant mutations. And this might make it unclear why the Stage 1 selection for veinlessness-if-heat-shocked is needed en route to innate veinlessness: if the genes for innate veinlessness were already available at Stage 0, then wouldn't innate veinlessness always have been open to selection over veinlessness-if-heat-shocked, given the artificial selective regime imposed by Waddington? However, the availability of genes does not automatically mean that they will occur together.. If both the first gene for lack of heat shock protein and the second gene for the deformed developmental protein and were rare in the original population, then there would have been little chance of their ever occurring together (or of their remaining together if they did). So the Stage 1 selection of veinlessness-if-heat-shocked is still essential, in order to provide a population where the first gene is common, and the second gene therefore likely to yield the phenotype innate veinlessness required for its selection.

With niche construction Baldwin effects, by contrast, the prior learning of P alters selection pressures so as render genes for P advantageous, where they wouldn't have been advantageous otherwise. I shall consider two ways in which niche construction can yield Baldwin effects. First, I shall briefly look at an idea of Peter Godfrey-Smith's, which I shall call 'keeping up with the Joneses'. Then, in the next section, I shall show that social learning is itself a form of niche construction which yields powerful Baldwin effects.

Niche construction itself extends far more generally than the Baldwin effect. It occurs whenever some new activity by some population creates new selection pressures on their genes. For example, the evolution of innate adult lactose tolerance in some human populations is a response to new selection pressures generated by the availability of milk from domesticated cattle. Again, the innate disposition of cuckoo chicks to eject host eggs from the nests is clearly a genetic adaptation to the parental cuckoo practice of parasitizing other species' nests. (Cf. Laland et al, 2000.) However, these examples are not Baldwin effects, as I am understanding the term, since they are not cases where the learning of some behaviour lead to the innateness of that self-same behaviour. (Rather, dairy farming leads to innate lactose tolerance; parental nest parasitizing leads to innate egg ejection by offspring.)

Godfrey-Smith, in considering niche construction as a possible source of Baldwin effects, focuses on the possibility that it may become more important to do P when everybody else is doing it (thus 'keeping up with the Joneses'). In such cases, the widespread learning of P through some population may itself increase the selective advantage of acquiring P quickly and reliably—that is, via genes rather than learning. In such cases, the selective coefficient of genes for P would display a kind of 'positive frequency dependence'—their selective advantage would increase as P becomes more widespread in the population. (Note, however, that this is not the kind of 'frequency dependent selection' normally discussed in the population genetics literature, in that here the selective coefficient will increase with the frequency of the phenotype P, which may be learned as well as innate, rather than the frequency of some allele that makes P innate.)

It is not obvious that the Jones's mechanism will work. Suppose that some trait P can either be genetically fixed by allele P_G, or left to learning by allele P_L. If P is adaptive, and P_G delivers it more quickly and reliably than P_L, as we are assuming, then won't P_G already have an advantage over P_L at Stage 0, even before any organisms acquire P from learning at Stage 1?

This is true enough. But it remains possible that P_G may have an even greater advantage over P_L at Stage 1, and that this increased advantage may be crucial in driving it to fixation. Imagine that the environment changes so that some crucial resource becomes too scarce for all to enjoy it—a climatic change means there are now only enough nuts for 90% of the population, say. Animals who are able to climb nut trees (P) are able to get nuts ahead of those who have to wait for the nuts to fall. But tree climbing is very laborious to acquire from learning, as opposed to getting it innately from some gene P_G. In such a case, P_G will indeed have some slight advantage over P_L even at Stage 0 when scarcely any animals can in fact climb trees, since it will eliminate any danger of ending up with no nuts, yet avoid the costs of

learning P. However, this advantage will not be great at Stage 0, for the animals will still have a good chance of getting nuts without climbing trees at all, and so a lack of P_G won't make it essential for them to incur the costs of learning P. At Stage 1, however, when most of the population has learned to climb trees, there will be no chance of getting nuts without climbing, and so any animal without P_G will be forced to undergo the costs of learning P in order to survive, thus greatly increasing the selective advantage of P_G over P_L.[9[9]]

So I agree with Godfrey-Smith that niche construction Baldwin effects might sometimes arise from biological imperatives to 'keep up with the Jones's'. However, by focussing on this kind of case, Godfrey-Smith seems to me to miss a far more obvious and important species of niche construction Baldwin effects, namely, those occasioned by social learning.

7 Social Learning as Niche Construction

Recall the kind of example I discussed in the Introduction. Some complex behaviour P is socially learned, where it is highly unlikely that any individual could learn P on its own. To vary our example, consider the common herring-gull practice of opening shellfish by grasping them in their beaks, flying up to a suitable height, dropping the shellfish on a hard surface, and retrieving the flesh from the broken shell. There is reason to suppose that this behaviour is socially transmitted. [Ref?] Now, once a given populations of gulls possesses this culture for opening shellfish, then this itself will create selection pressures for genes that make them better at acquiring it. An individual with an allele that innately disposes it to grasp clams when it sees them, say, will learn how to get shellfish meat more quickly, since it will have less to learn than gulls who lack this allele. But note that this allele would have no selective advantage, were it not for the pre-existing culture of shellfish-dropping, given that there would be no real possibility of learning the rest of the complex behaviour without any exemplars to copy from. There's no advantage in being disposed to grasp clams when you find them, if you don't then fly up, drop the clams, and retrieve the meat—and even gulls for whom the grasping disposition is innate would be highly unlikely to figure out the rest of this behaviour by individual trial-and-error if they could not learn it from other gulls.

So this then gives us another kind of niche construction Baldwin effect. The prior existence of some learned behaviour (Stage 1) creates selection pressures for genes that will render that selfsame behaviour innate. And the prior learning of that behaviour is indeed essential here, since the genes in question would have no selective advantage in an environment (Stage 0) where no animals were learning the behaviour and so providing exemplars for further learners.

The analysis of such social learning Baldwin effects is complicated by the fact they will inevitably involve genetic assimilation as well as niche construction. To see this, note that, when a socially learned behaviour creates selection pressures for genes for components of that behaviour, it is precisely by making the behaviour easier to learn.

[9[9]] Note that this mechanism doesn't require that there is selection of genes for learning P at Stage 1, as originally required by Godfrey-Smith (2003). It is quite enough that P is produced by general and long-standing learning mechanisms operating in some new environment. (But see Godfrey-Smith, forthcoming, where he corrects this.)

The social learned behaviour is significant as a niche constructor because at earlier stages it enables the remaining components of the overall behaviour to be learned.[10[10]]

So, insofar as some socially learned behaviour functions as an environmental niche that selects for its own innateness, it will be by lightening learning loads, which means that the requirements for genetic assimilation will also be satisfied. As in section 4, we will have a complex behaviour with a number of components, none of which is adaptive on its own. Given this, genes for those components might individually seem to lack any selective advantage, given the improbability of any one of them finding itself together with the others. However, once we take learning into account, then we can see that these genes are individually advantageous after all, since each on its own lightens the amount of learning needed to acquire the overall adaptive P.

Still, it would be a mistake to think that the niche construction aspect of social learning adds nothing to the genetic assimilation mechanism discussed earlier. The niche construction aspect also tells us why it is possible for organisms to learn all the rest of P when only a few components of the behaviour are innate. Earlier, when discussing genetic assimilation itself, we simply took it for granted that such learning would be possible. However, in the cases now at issue, it is highly unlikely that any animals with only a few relevant genes will be able to learn all the rest of P by individual trial-and-error—in our example, it was highly unlikely that even a herring gull for whom clam-grasping is innate would be able to learn all the rest of the clam-opening behaviour on its own. The prior learned culture of P is thus essential for an environment in which the rest of P can be socially learned. It is precisely because the other gulls are already displaying the clam-opening behaviour that a tyro with only a few innate elements can acquire the rest of the behaviour.

With socially learned behaviours, then, we get Baldwin effects twice over. The prior learning of P (Stage 1) is crucial to P's becoming innate in two quite different ways. Not only does P need to be learned while each of the earlier I_Gs get selected, as in all genetic assimilation, but also the niche construction means there wouldn't be any selective pressure on those I_Gs to start with unless the socially learned P were being displayed by conspecifics.

8 The Significance of Social Learning

I have focused on the structure of two kinds of Baldwin effect: genetic assimilation and niche construction. And I have argued that the genetic selection of social learned behaviours can constitute both kinds of Baldwin effect simultaneously. But is this anything more than a conceptual oddity? Why should it matter that certain possible

---

[10[10]] Thus note how the niche construction story ceases to apply at the point where P becomes entirely innate. Imagine that genes for all but the 'last' component P_n in some complex behaviour have already become fixed in the population, and consider the remaining competition between alleles N_G and N_L for this last component. This last N_G will still have an advantage over its competing N_L, since it ensures P more cheaply and reliably. However, this last advantage won't depend on the prior culture of P, since once an animal has this last N_G it will have nothing left to learn from its conspecifics. Given that the other genes are all in place, this last N_G would be favoured over the alternative N_L even if no other animals were displaying P.

processes may fit the half-formed ideas of an unimportant nineteenth-century theorist in two different ways?

Well, there's nothing especially significant about possible Baldwin mechanisms, even doubly Baldwinian ones. To show that certain processes are in principle biologically possible is of merely theoretical interest, in the absence of any reason to think that they are empirically important. Maybe the social learning of certain behaviours can lead to their own innateness in a way that fits Baldwin's conjecture twice over. But unless such cases play an important empirical role in evolution, this would be nothing more than an odd quirk of intellectual history.

However, I think that there is some reason to suppose that these doubly Baldwinian social learning processes are empirically important. This is because social learning vastly expands the class of learnable adaptive behaviours. Many behaviours are far too complex for animals to have any realistic chance of acquiring them by individual learning alone (even with one or two genes to help them on their way). So, if everything was left to individual learning, these behaviours could never be genetically assimilated—with no real chance of the rest of the behaviour being learned, no I_G would have any initial selective advantage, and genetic assimilation wouldn't get going. But now add in social learning. This means that as soon as one lucky or exceptional individual somehow acquires P, then it becomes possible for the others to pick up P socially, when it wouldn't possible for them to learn P otherwise.[11[11]] And this will give the relevant I_Gs an initial advantage after all, and allow genetic assimilation to get going. The point is that genetic assimilation requires that the relevant P be learnable, and social learning renders many interesting Ps learnable when they wouldn't otherwise be.

Because of this, I suspect that precisely my double Baldwin effect is responsible for the innateness of many complex behavioural traits. If we have genetic assimilation, then that is one kind of Baldwin effect: the components of some adaptive behaviour progressively come under genetic control, because each relevant gene facilitates learning the rest of that same behaviour. However, in many such cases the relevant genes wouldn't facilitate learning the rest of that behaviour, were it not for the help of the second kind of Baldwin effect: some adaptive behaviour is learnable, and so open to genetic assimilation, only because that same behaviour is available as an exemplar for social learning.

Of course, this process does require that 'one lucky or exceptional individual somehow acquires P' independently of social learning, in order to get the social promulgation of P off the ground. And this prerequisite may seem to be in some tension with the idea that double Baldwin effects will be important precisely with Ps that are 'too complex for animals to have any realistic chance of acquiring them by individual learning alone'. But this tension is more apparent than real. For note that,

---

[11[11]] Here is one point where differences between different kinds of social learning may matter (cf. section 2.3 above). Suppose that some unusual individual does acquire some adaptive behaviour P non-socially. What ensures that others will copy this individual, rather than others without P? Not all modes of social learning would seem to privilege models who display adaptive behaviours over others—in particular, blind mimicry—(iii)—won't do this. But perhaps that doesn't matter, if we suppose that individual reinforcement acts as a moderator of social learning, perpetuating only those behaviours that yield reinforcing rewards.

in the absence of social learning, all individuals would need to be able to acquire P by individual learning, in order for genetic assimilation to occur. Given social learning, however, only one individual need acquire P non-socially, in order to get things moving. There is no reason why one such lucky strike shouldn't be reasonably probable, even if the chance of any given individual acquiring P non-socially is very low. (If the probability of success in a single trail is p, the probability of at least one success in K independent trials is $(1-(1-p)^K)$.)[12[12]]

In my Introduction I said that I would defend the importance of Baldwin effects against those who say that they are at best special cases of the more general phenomena of genetic assimilation and niche construction. I have now argued that, when the social learning of some behaviour leads to its own innateness, genetic assimilation and niche construction combine to produce a particularly powerful mechanism of natural selection. I take this to show that this kind of Baldwin effect at least is worth singling out for special attention.

Let me conclude with an empirical prediction. If my doubly Baldwinian social learning mechanism has indeed been important for the evolution of complex behaviours, as I have hypothesized, then we should expect to find, somewhat paradoxically, that complex innate behaviours are more common in species that are good social learners than in other species. True, any such correlation will be diluted by the fact that the relative costs of learning and genetic control will not always favour bringing socially learned traits under genetic control (in the way I have been assuming since section 2.1). Even so, if I am right in thinking that social learning vastly expands the range of behaviours open to genetic assimilation, species that are good general social learners should still display significantly more complex innate behaviour than other species. Unfortunately I lack the expertise to assess this prediction myself. But I would be very interested indeed to know whether or not the comparative zoological data bear it out.[13[13]]

Bibliography

Akins, C. and Zentall, T. 1998. 'Imitation in Japanese quail: the role of reinforcement of demonstrator responding'. Psychonomic Bulletin and Review, 5, 694-7.

Baldwin, J. M. 1896. 'A New Factor in Evolution', The American Naturalist 30, 441-51, 536-53.

Boyd, R. and Richerson, P. 1996. 'Why Culture is Common but Cultural Evolution is Rare', Proceedings of the British Academy 88, 73–93.

[12[12]] Following Godfrey-Smith, I have specified that a Baldwin effect begins with a Stage 0 where 'the environment changes so as to make phenotype P adaptive'. But perhaps such environmental shifts are not always necessary: in some cases the Baldwinization may begin simply because some individual animal happenstantially acquires P and the practice then spreads socially.

[13[13]] I would like to thank Nell Boase, Paul Griffiths, Peter Godfrey-Smith, Paul Rozin, Tom Simpson, Stephen Stich, Elliott Sober, Kim Sterelny, and especially Matteo Mameli for comments on previous versions of this paper.

Cowie, F. 1999. What's Within? Nativism Reconsidered. , New York: Oxford University Press.

Downes, S. 2003. 'Baldwin Effects and the Expansion of the Explanatory Repertoire in Evolutionary Biology', in Weber, B. and Depew, D., Learning, Meaning and Emergence: Possible Baldwinian Mechanisms in the Co-Evolution of Mind and Language, [/]

Godfrey-Smith, P. 2003. 'Between Baldwin Scepticism and Baldwin Boosterism', in Weber, B. and Depew, D., Learning, Meaning and Emergence: Possible Baldwinian Mechanisms in the Co-Evolution of Mind and Language, [/]

Godfrey-Smith, P. Forthcoming. 'On the Evolution of Representative and interpretational Capacities.' Monist.

Griffiths, P. 2003. 'Beyond the Baldwin Effect: James Mark Baldwin's 'social heredity', epigenetic inheritance and niche-construction', in Weber, B. and Depew, D., Learning, Meaning and Emergence: Possible Baldwinian Mechanisms in the Co-Evolution of Mind and Language, [/]

Griffiths, P.E (2002) 'What is Innateness?' The Monist 85, 70-85

Hinton, G. and Nowlan, S. 1987. 'How Learning can Guide Evolution.' Complex Systems 1, 495-502.

Hunt, G. and Gray, R. 'Diversification and Continual Evolution in New Caledonian Crow Toll Manufacture'. Procedings of the Royal Society of London B, **739, 1-9**.

Laland, K., Olding-Smee, J., and Feldman, M. 2000. 'Niche Constrcution, Biological Evolution, and Cultural Change'. Behavioural and Brain Sciences 23, 131-75.

Mayley, G. 1996. 'Landscapes, Learning Costs, and Genetic Assimiliation', in Turney, P., Whitely, D., and Anderson, R. (eds) Evolutionary Competition, Evolution, Learning and Instinct: 100 Years of the Baldwin Effect.

Samuels R. (2002) 'Nativism in Cognitive Science'. Mind and Language 17, 233-265

Shettleworth, S. 1998. Cognition, Evolution and Behavior. Oxford: Oxford University Press.

Simpson, G. G. 1953. 'The Baldwin Effect', Evolution 7, 110-17.

Sterelny, K. 2000. The Evolution of Agency and Other Essays. Cambridge: Cambridge University Press.

Tebbich, S., Taborsky, M., Fessl, B., and Blomqvist D. 2001. Do woodpecker finches acquire tools use by social learning? Proceedings of the Royal Society 268: 2189-2193.

Tomasello, M. 2000. The Cultural Origins of Human Cognition. Cambridge, Mass: Harvard University Press.

Waddington, C.H. 1953. ‘Genetic Assimilation of an Acquired Character’. Evolution 4, 118-26.

Watkins, J. 1999. ‘A Note on Baldwin Effect’, British Journal for the Philosophy of Science 50, 417-23.

Whiten, A., Goodall, J., McGrew, W., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C., Wrangham, R., Boes, C., and Boes **?** 1999, ‘Culture in Chimpanzees’. Nature 399, 682-5.