Title

What the Baldwin Effect affects

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors

Morgan, Thomas J.H.
Griffiths, Thomas L

Publication Date

2015

Peer reviewed

What the Baldwin Effect affects

Thomas J. H. Morgan (thomas.morgan@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California, Berkeley, CA 94720 USA

Abstract

The Baldwin Effect is a proposed mechanism by which plasticity facilitates adaptive phenotypic and genetic evolution. In particular it has been proposed to be involved in the evolution of language. Here we investigate three factors affecting the extent to which plastic traits are fixed by selection: (*i*) the difficulty with which traits can be acquired through plasticity, (*ii*) the importance of traits to fitness, and (*iii*) the nature of dependencies between different traits. We find that selection preferentially fixes traits that are difficult to acquire through plasticity, traits that have larger fitness benefits, and traits that affect the acquisition of, or benefits from, other traits. We conclude by discussing the implications of these findings for the evolution of language as well as non-human behaviors and reconsider the evolutionary significance of the Baldwin Effect.

Keywords: Baldwin effect; gene-culture co-evolution; language evolution.

Introduction

The relationship between phenotypic and genetic change is an outstanding question in evolution. While the typical view is that phenotypic change is the result of genetic change, this contrasts with observations of living organisms where plasticity (the capacity to respond to environmental inputs with phenotypic change) often produces more rapid phenotypic change than would be possible via genetic change (West-Eberhard, 2003).

One proposal for how development and evolution can interact to shape the phenotype is the "Baldwin Effect", a term coined by G. G. Simpson in 1953 in reference to J. M. Baldwin's 1896 publication "A New Factor in Evolution". It refers to the process by which selection reduces the plasticity of traits that initially arose via plasticity. Models of the Baldwin Effect have shown that plasticity can greatly accelerate both phenotypic and genetic evolution (Hinton & Nowlan, 1987). This is because plasticity can expose variation in the ability to acquire a trait during development, where without such plasticity there would be no relevant variation.

The Baldwin Effect has also been invoked in theories of the evolution of language (e.g., Deacon, 1997); language's universal distribution is intuitively suggestive of genetic influence, whilst the variation between languages and the time required to learn a language show it is also plastic. Whilst some authors (Pinker & Bloom, 1990; Pinker, 2003) have used the Baldwin Effect to argue for the evolution of a strong genetic influence specific to language, other work suggests the Baldwin Effect is unlikely to produce such specific results (Chater et al., 2009; Christiansen et al., 2011).

These differences highlight how, despite recent interest, basic questions about the Baldwin Effect remain unanswered. A key question concerns what kinds of traits the Baldwin Effect is likely to affect. If the evolution of language involved the Baldwin Effect, what aspects of language should we expect to see under relatively strong genetic influence? We address this question by using evolutionary simulations to explore three factors that affect the extent to which selection reduces the plasticity of traits in a stationary environment. The factors we consider are: (*i*) the ease with which traits can be acquired through plasticity, (*ii*) the fitness contributions of traits, and (*iii*) dependencies concerning the acquisition of plastic traits. We find that selection preferentially fixes traits that are hard to acquire through plasticity, traits that are are important to fitness, and traits upon which the acquisition of other plastic traits depend. We also show that the extent of fixation is negatively affected by the number of traits to be fixed and the mutation rate.

Background

At the end of the 19th century, after the publication of *The Origin of Species* (Darwin, 1859), but prior to the formation of the Modern Synthesis, there was still considerable debate over the mechanism of evolution (Larson, 2004). Lamarckian proposals, involving the inheritance of acquired characteristics, were falling out of favor, due to the vigorous argumentation of August Weissmann (Haig, 2007), and also the failure of Lamarckians to produce positive evidence of the inheritance of acquired characteristics (Larson, 2004). Meanwhile, various developmentally inclined researchers proposed what was then known as "organic selection"; the idea that developmental plasticity, by adaptively shaping phenotypic variation, directs, and thus predicts, genetic change without the need for the inheritance of acquired characteristics (Baldwin, 1896; Depew, 2003; Osborne, 1896; Lloyd Morgan, 1896; Godfrey-Smith, 2003). However, the implications of organic selection were never established and an increasing focus on genes meant that development was not included the Modern Synthesis (Amundson, 2005). It was not until the 1950s that organic selection was re-named the Baldwin Effect (Simpson, 1953), but the general conclusion was that, whilst possible, the Baldwin Effect lacked any real importance. Soon after, the Baldwin Effect, having become associated with Lamarckian ideas, was virtually abandoned (Mayr, 1963).

However, the past 30 years have seen a modest resurgence of interest in the Baldwin Effect. This was, in part, due to theoretical work showing that learning can accelerate the evolution of difficult-to-find traits (Hinton & Nowlan, 1987). Hinton and Nowlan first considered a trait underpinned by multiple genetic loci, each of which could have two alleles: *correct* or *incorrect*. The only way for the organism to gain a fitness benefit was to possess all *correct* alleles. As the number of loci increases, the odds of finding such a needle in

a haystack diminish tremendously and populations are fated to drift randomly amongst functionally equivalent genotypes until the single effective genotype is stumbled upon. This was contrasted with a scenario including a third, plastic, allele, *?*. Provided an organism's genome contains only *correct* and *?* alleles, they at least have a chance at discovering the trait, an outcome that is increasingly likely as the number of *correct* alleles increases. The inclusion of plasticity greatly accelerated the evolution of the beneficial trait, due to two factors: (*i*) there are now many possible genotypes capable of learning the trait, not just one, ameliorating the needle-in-a-haystack problem, and (*ii*) once an organism capable of learning the trait evolves there is a continuous selection gradient towards more effective learning by increasing numbers of *correct* alleles.

Several authors have implicated the Baldwin Effect in accounts of the evolution of language (Pinker & Bloom, 1990; Pinker, 2003; Deacon, 1997). One such approach is the "cognitive niche" theory of human evolution (Pinker, 2010, 2003; Pinker & Bloom, 1990; Barrett et al., 2007). In this account, a genetic language capacity coevolved with other cognitive abilities in the face of an environment that favored a complex and flexible "improvisational intelligence". A difficulty, however, lies in how genes supporting language could spread from a single individual, given that their benefit is reliant on them being widespread in a population. This problem is analogous to that considered by Hinton and Nowlan (1987), but rather than coordinating multiple loci within one individual, the problem is to coordinate the same locus across individuals. Furthermore, the cost of sustaining a brain capable of language would seemingly prevent the mutation from spreading via drift. Pinker and Bloom (1990) argued that this problem could be solved with the Baldwin Effect. If all individuals have some capacity to learn language through plasticity, but the mutation greatly enhances this capacity, when the mutation arises its bearer will be able to share language through social interactions. This allows the mutation to increase fitness even if it is unique in the population and so it can spread through natural selection. However, whilst plausible, the cognitive niche has been criticized for taking insufficient account of the role of culture in the evolution of language (Boyd et al., 2011; Whiten & Erdal, 2012; Heyes, 2012).

An alternative account, closer to late 19$^{\text{th}}$ century ideas of plasticity not only accelerating, but also guiding evolution, has been proposed by Deacon (1997). In this account, language was initially a product of cultural evolution, drawing on more general cognitive abilities. The benefits from language, however, generated selection favoring organisms who could learn the language more easily. Accordingly, selection favored genetic change that enhanced the ability to acquire the language, eventually resulting in modern human language capabilities. This account fits well with theories such as niche construction (Odling-Smee et al., 2003) which explores the evolutionary consequences of organisms abilities to modify their selective environment, and gene-culture co-evolution (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981; Lumsden & Wilson, 1981; Richerson & Boyd, 2005) which studies interactions between genetic and cultural evolution.

Given the above theorizing, a small number of papers have used mathematical models to explore the extent to which selection can fix plastic traits. Ancel (1999; 2000) found that whilst plasticity can accelerate the initial appearance of a new trait, it actually slows it's fixation because, once the trait can be reliably acquired, there is little selective advantage to its fixation. Other work has argued that only highly stable plastic traits can be fixed (Chater et al., 2009) as otherwise the traits will change before selection on genes can fix them. Initially Chater et al. (2009) suggested that language changes too rapidly for fixation to be plausible, however, in later work they suggest that the stable features shared across languages could be subject to the Baldwin Effect, but emphasize that the fixed traits may not be language specific (Christiansen et al., 2011). Supporting this, other work has found that the Baldwin Effect can be prevented if the resulting changes would result in cognitive constraints that decreased performance at other tasks (Reali & Christiansen, 2009). The involvement of the Baldwin Effect in the evolution of language thus remains a real possibility, though the evolution of a *language-specific* genetic capacity perhaps less so.

Whilst existing work reveals the circumstances under which the Baldwin Effect can operate, an orthogonal set of questions concern which traits are likely to be affected. In the remainder of the paper we use a simulation approach to consider the fixation of a group of traits of different (*i*) difficulty to acquire through plasticity, (*ii*) fitness importance and (*iii*) interdependency. We find that selection preferentially fixes difficult traits, traits that are significant to fitness and traits upon which others depend. We also find that plasticity is most prevalent when there are a large number of traits. The end result is that complex behaviors consisting of multiple steps can be performed with deceptive ease and effectiveness, despite still requiring learning. We relate these findings to the evolution of language and other traits more generally.

The Simulation Framework

Our simulation framework is based on that of Hinton and Nowlan (1987). We consider a population of N asexual, haploid organisms whose fitness is determined by the acquisition of n fitness-relevant traits. The successful acquisition of the t^{th} trait by the i^{th} individual is affected by a corresponding genetic locus, $G_{i,t}$. Accordingly, each organism has n genetic loci. Each locus contains one of two possible alleles: *fixed* and *plastic*. A *fixed* allele means that the organism is guaranteed to acquire that trait, whilst a *plastic* allele means that it is capable of acquiring the trait through plasticity with probability p. Hinton and Nowlan (1987) included a third allele that prevented individuals from acquiring the relevant trait, however, this was included to show that the existence of *plastic* alleles accelerates the rate at which the trait can be found. Here our intent is to explore factors affecting the extent of fix-

ation at different loci assuming that the traits can be acquired with a non-zero probability, and so we do not include the third allele. The probability that the i^{th} individual acquires the t^{th} trait (whether through plasticity or fixation) is is given by $\phi_{i,t}$, where:

$$\phi_{i,t} = \begin{cases} 1 & \text{if } G_{i,t} = \textit{fixed} \\ p_{i,t} & \text{if } G_{i,t} = \textit{plastic} \end{cases} \quad (1)$$

The fitness benefit to the i^{th} individual from the t^{th} trait is $f_{i,t}$. Accordingly, the fitness, F, of the i^{th} organism is given by:

$$F_i = F_{min} + \sum_{t=1}^{n} \phi_{i,t} f_{i,t} \quad (2)$$

where F_{min} is a baseline value of fitness, assumed to come from other fitness relevant traits (whilst F_{min} simply serves to avoid the floor effects of organisms having negative fitnesses, higher values of F_{min} will also weaken selection). Once fitness is calculated, organisms reproduce according to the Wright-Fisher process where repeated sampling with replacement from the parental generation, weighted by fitness, generates N offspring (Wright, 1931; Fisher, 1930). The offspring generation then replaces the parental generation. Every offspring inherits their parent's genome subject to mutation. Each locus mutates, changing its allele, with probability q.

Unless specified otherwise: $p_{i,t} = 0.75, f_{i,t} = 1, F_{min} = 1$ and $N = 500$. Simulations were carried out for values of n between 2 and 20 and values of q between 0.001 and 0.1 and starting genotypes were a random assortment of *fixed* and *plastic* alleles.

Simulation 1: Difficulty

First, we explore how differences in the probability traits are acquired through plasticity affect the extent to which they become fixed. To do this, we assume all behaviors give all individuals the same benefit (i.e., $f_{i,t}$ is the same for all i and all t) and that all individuals are equally effective at acquiring traits through plasticity (i.e., $p_{i,t}$ is the same for all i), but we do not assume that all behaviors are equally easy to acquire through plasticity (i.e., $p_{i,t}$ is not the same for all t). Specifically, we considered the case where, for a given value of n, the probability that the t^{th} trait is acquired through plasticity is given by:

$$p_{i,t} = \frac{1}{2} + \frac{t - \frac{1}{2}}{2n} \quad (3)$$

This produces a range of values evenly spaced between 0.5 and 1, with lower values of t associated with lower values of $p_{i,t}$. For robustness, we carried out simulations with other distributions of $p_{i,t}$ (including ranging from 0 to 1), but results were always consistent.

The simulation results reveal that the more difficult a trait is to acquire through plasticity, the greater the extent to which selection fixes it (see Figure 1). Assuming $q = 0.01$, the easiest trait to acquire through plasticity is between 10 and 20% more plastic than the most difficult trait. Average fitness is higher for a given level of average plasticity than would be expected if plasticity were evenly distributed across loci because the more plastic traits are relatively easy to acquire through plasticity. The extent of fixation is reduced by increasing the number of traits; assuming $q = 0.01$ average fixation is ~85% for $n = 2$, but ~60% for $n = 20$. This is because, as each locus has a chance of mutating independent of all other loci, the expected number of mutations per offspring, $\hat{q}$, is given by:

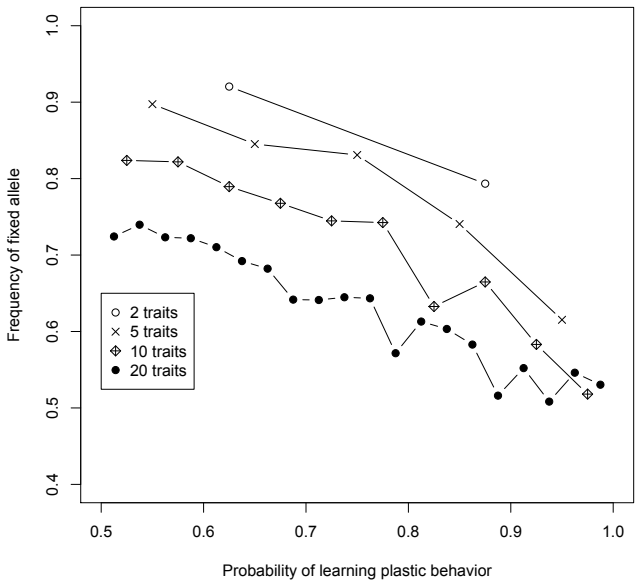

Figure 1: The extent to which selection fixes traits is affected by how hard they are to acquire through plasticity ($p_{i,t}$). This figure shows the fixation of each trait, for n = 2, 5 10 and 20. Each point shown is the average of 20 repeats of 500 generations, with $N = 500$ and $q = 0.01$.

$$\hat{q} = nq \quad (4)$$

Accordingly, the ability of selection to fix traits is reduced when there are more traits to fix or when the probability of a locus mutating is increased. To test robustness, further simulations were run with n = 50, 100 and 200 and the relationship between difficulty and fixation remained, although, as expected, overall fixation was lower. It should be noted however, that the value of q we used is unrealistically high and using more realistic values increases the extent of fixation.

Simulation 2: Importance

We now explore how differences in the fitness effects of traits affect the extent to which they become fixed. To do this, we assume that all individuals are equally effective at acquiring traits through plasticity (i.e., $p_{i,t}$ is the same for all i) and that all traits are equally easy to be acquired through plasticity

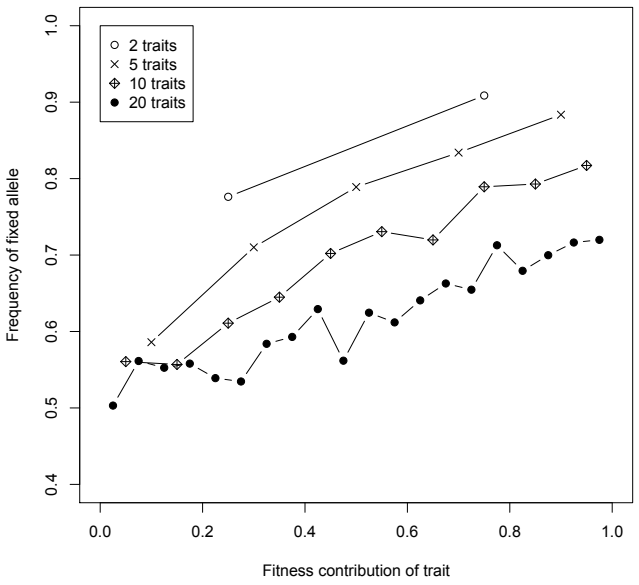

Figure 2: Selection preferentially fixes important traits (those with larger values of $f_{i,t}$). This figure shows the plasticity of each trait for populations with 2, 5, 10 and 20 traits after 500 generations. Each point shown is the average of 20 repeats, with N=500 and q=0.01. The probability of each trait being acquired if plastic ($p_{i,t}$) is 0.75.

(i.e., $p_{i,t}$ is the same for all t). However, we assume different traits bring different fitness benefits (i.e. $f_{i,t}$ is the same for all i, but not all t). Specifically we considered the case where, for a given value of n, the fitness benefit of acquiring the t^{th} trait is given by:

$$f_{i,t} = \frac{t - \frac{1}{2}}{n} \tag{5}$$

This produces a range of values evenly spaced between 0 and 1, with lower values of t associated with lower values of $f_{i,t}$.

The results show that the more important a trait is to fitness, the greater the extent to which it is fixed by selection (see Figure 2). Assuming q = 0.01, the trait with the greatest effect on fitness is between 10 and 20% less plastic that the least important trait. This is because fixing important traits brings more fitness benefit than fixing less important traits and so selection acts to genetically secure the more valuable traits, whilst leaving the less valuable ones up to plasticity. As before, average fitness is greater than if plasticity were equally distributed across loci and increasing the number of loci or the mutation rate increases average plasticity.

Simulation 3: Dependence

Finally, we consider a case where the traits depend on each other. Specifically, we assume that the t^{th} trait depends on traits 1 to (t-1). One way such dependence could be manifest is where the *acquisition* of the t^{th} trait depends on the acquisition of traits 1 to (t-1). Taking this dependency into account, the probability that the i^{th} individual acquires the t^{th} trait, $\phi_{i,t}$, is:

$$\phi_{i,t} = \begin{cases} 1 & \text{if } G_{i,t} = \textit{fixed} \\ p_{i,t} \prod_{u=1}^{t} \phi_{i,u} & \text{if } G_{i,t} = \textit{plastic} \end{cases} \tag{6}$$

Accordingly, the fitness of the i^{th} individual is:

$$F_i = F_{min} + \sum_{j=1}^{n} \left(f_{i,j} \prod_{k=1}^{j} \phi_{i,k} \right) \tag{7}$$

An alternative way to imagine this scenario is that the fitness effect of acquiring the t^{th} trait is contingent on the acquisition of traits 1 to (t-1). Taking this dependency into account, the probability that the i^{th} individual acquires the t^{th} trait, $\phi_{i,t}$, is given by equation 1, whilst the fitness effect of the t^{th} trait to the i^{th} individual is:

$$f'_{i,t} = f_{i,t} \prod_{u=1}^{t} \phi_{i,u} \tag{8}$$

where $f_{i,t}$ is the fitness effect of trait acquisition, assuming all traits upon which it depends have been acquired (either through plasticity or fixation). In this case the fitness of the i^{th} individual is:

$$F_i = F_{min} + \sum_{j=1}^{n} \left(f_{i,j} \prod_{k=1}^{j} \phi_{i,k} \right) \tag{9}$$

Equations 7 and 9 are the same, and so both interpretations will have the same evolutionary consequences and are interchangeable.

Note that, aside from effects of dependency, we assume that all traits are equally difficult to acquire through plasticity and bring the same fitness benefit (i.e., both $p_{i,t}$ and $f_{i,t}$ are the same for all i and all t).

The results show that the larger the number of traits that depend upon a particular trait, the greater the extent to which selection fixes it (see Figure 3). Assuming q=0.01, the most foundational trait is between 10 and 20% less plastic than the most trivial trait. This is because of the influence foundational traits have on the acquisition/fitness effects of other traits, their fixation brings larger increases to fitness. As before, average fitness is greater than if plasticity were equally distributed across loci and increasing the number of loci or the mutation rate increases average plasticity.

Discussion

We have shown that the genetic fixation of plastic traits – the Baldwin Effect *sensu* Simpson (1953) – is not indiscriminate and that selection preferentially fixes traits that are difficult to acquire through plasticity, that come with larger benefits to fitness, or that affect the probability of acquiring other traits

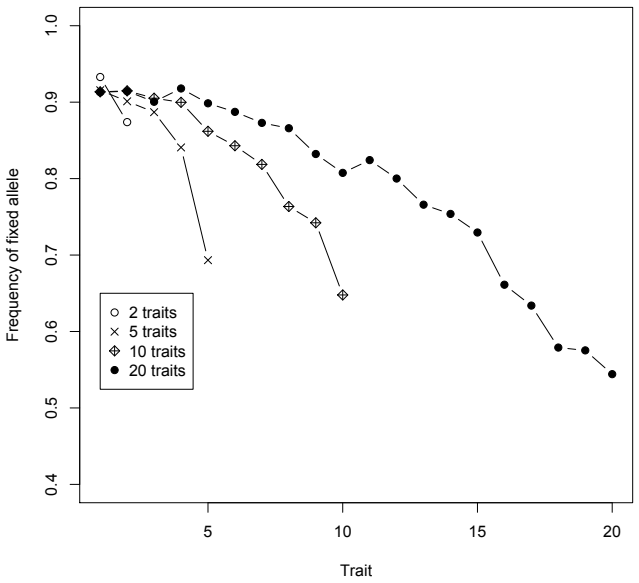

Figure 3: Selection preferentially fixes traits that are needed to acquire other traits through plasticity or to gain the fitness benefits from other traits. This figure shows the plasticity of each trait for populations with 2, 5, 10 and 20 traits after 500 generations. The horizontal axis shows the indices of the traits, where each trait can only be acquired via plasticity once all traits with lower indices have been acquired. Accordingly, for the simulations with only two traits, there are only two points at indices 1 and 2. Each point shown is the average of 20 repeats, with $N = 500$, $q = 0.01$ and $p_{base} = 0.75$.

(or their fitness benefits). In all three simulations the equilibrium populations still show extensive plasticity, and individuals typically do not acquire all the traits, yet they acquire more traits and have higher fitness than would be expected if the plasticity were distributed evenly across traits. This is because the acquisition of the difficult, important or integral traits has been enhanced by selection. Together with the work of Chater et al. (2009) this identifies a potential feedback mechanism by which traits that are sufficiently stable to be subject to the Baldwin Effect become even more stable as their acquisition is made more reliable.

These findings are not examples of gene-culture coevolution or niche construction (unless plasticity itself is considered an example of niche construction) as there is no cultural transmission or external modification of selective environments within the models. Rather, they illustrate the interaction between developmental plasticity and selection. We do not mean that cultural or niche constructing accounts of the Baldwin Effect are misguided, just that the Baldwin Effect is a more general process that also includes non-cultural and non-niche constructing cases. Nonetheless, it remains likely that culture greatly increases the importance of the Baldwin Effect by bringing an increasing number of traits within the scope of plasticity - something that Baldwin discusses as "social heredity" (Baldwin, 1896). Indeed, such a suggestion can be seen in both Pinker and Bloom's (1990) and Deacon's (1997) use of the Baldwin Effect to explain the evolution of language. In both cases communication between individuals, culturally transmitting the ability to engage in language, is necessary for the genetic evolution of an increased language capacity. The difference between the two is that for Pinker and Bloom (1990) the emergent language is specified by a genetic language ability, whereas for Deacon (1997) the emergent language is a product of cultural evolution that becomes genetically entrenched.

Our findings suggest that if a multi-faceted trait (such as language) were subject to the Baldwin Effect, even at equilibrium, its acquisition should require learning and so errors in acquisition will occur (because plasticity is not extinguished), but, despite this, its complex aspects should be remarkably easy to learn and its important or essential aspects should be widespread. This offers a potential answer to the question of how it is that language is so readily acquired by human children: the acquisition of the (seemingly) difficult, important or foundational aspects may have been enhanced by the Baldwin Effect. This raises the question of which aspects of language these are. Whilst we can only provide a tentative suggestion, features such as the existence of a grammar or tenses (as opposed to a given language's specific grammar or tense system) are foundational to the use of language and so satisfy our conditions for fixation (although how difficult they were for our ancestors to acquire is hard to say), and in keeping with our findings they are also common across languages. However, rather than aspects of language itself being fixed, a prosocial motivation to engage with others and socially acquire information could have been subject to the Baldwin Effect and indirectly fostered the cultural evolution of language. This is in line with work suggesting that any fixation would tend to favor the acquisition of many behaviors, as opposed to just language (Christiansen et al., 2011), and a tendency to be motivated to learn from others seems in line with human behavior which is dependent on culture to an unparalleled extent.

As plasticity is a ubiquitous feature of living organisms, our findings may have implications for non-human evolution. Two cases we will briefly discuss here are (*i*) the ability of archerfish to shoot down flying insects with jets of water propelled from their mouth (Schuster et al., 2006), and (*ii*) the singing ability of songbirds, such as zebra finches (Feher et al., 2014). In both cases the behavior is complex and plastic, yet both are performed readily and with a high degree of proficiency. In the case of archerfish an astounding level of performance can be achieved through a small number of observations of another individual and without practice (Schuster et al., 2006). In the zebra finch, even playback of a youngster's own initial attempts at song is sufficient for them to

develop complex song typical of adults (Feher et al., 2014). Such behaviors seem plausible cases of the Baldwin Effect leading to the fixation of behaviors that once required much more learning and future work could investigate this in more detail.

Despite a mixed history, there is now a growing body of evidence to suggest that the fixation of learnt behaviors can occur and several factors have been identified that affect its extent - here we propose difficulty, importance, and dependence as well as the number of loci. Given this there might be reason to think that the Baldwin Effect could be valuable after all, both explaining the evolution of traits that rely on coordinating the behavior of multiple individuals, or complex traits that are performed with high reliability. While we agree that the Baldwin Effect does offer such a tool, we would say that it is likely to be a much more general process. Even if traits can be understood in terms of the accumulated effects of mutations on the phenotype without the need for plasticity, does such a process seem likely given how plastic organisms are so many ways? From this perspective the Baldwin Effect is not so much a "new factor" in evolution (Baldwin, 1896) as it is simply the standard evolutionary process.

Acknowledgments. This work was supported by grant number BCS-1456709 from the National Science Foundation.

References

Amundson, R. (2005). *The Changing Role of the Embryo in Evolutionary Thought*. Cambridge, UK: Cambridge University Press.

Ancel, L. W. (1999, January). A quantitative model of the Simpson-Baldwin Effect. *Journal of theoretical biology*, *196*(2), 197–209.

Ancel, L. W. (2000, December). Undermining the Baldwin expediting effect: does phenotypic plasticity accelerate evolution? *Theoretical population biology*, *58*(4), 307–19.

Baldwin, J. (1896). A New Factor in Evolution. *The American Naturalist*, *30*(354), 441–451.

Barrett, H. C., Cosmides, L., & Tooby, J. (2007). The Hominid Entry into the Cognitive Niche. In S. W. Gangestad & J. A. Simpson (Eds.), *The evolution of mind: Fundamental questions and controversies* (pp. 241–248). New York, NY: The Guilford Press.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R., Richerson, P. J., & Henrich, J. (2011, June). The cultural niche: why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108 Suppl*, 10918–25.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.

Chater, N., Reali, F., & Christiansen, M. H. (2009, January). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(4), 1015–20.

Christiansen, M. H., Reali, F., & Chater, N. (2011, April). Biological adaptations for functional features of language in the face of cultural evolution. *Human biology*, *83*(2), 247–59.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: Murray.

Deacon, T. W. (1997). *The Symbolic Species: The co-evolution of language and the brain*. New York, NY: W. W. Norton.

Depew, D. J. (2003). Baldwin and His Many Effects. In B. H. Weber & D. J. Depew (Eds.), *Evolution and learning: The baldwin effect reconsidered.* Cambridge, MA: MIT press.

Feher, O., Suzuki, K., Okanoya, K., Ljubicic, I., & Tchernichovski, O. (2014). Birds tutored with their own developing song produce wildtype-like song as adults. *The Evolution of Language. Proceedings of the 10th International Conference on the Evolution of Language*, 433–434.

Fisher, R. (1930). *The genetical theory of natural selection*. Oxford: Clarendon.

Godfrey-Smith, P. (2003). Between Baldwin Skepticism and Baldwin Boosterism. In B. H. Weber & D. J. Depew (Eds.), *Evolution and learning: The baldwin effect reconsidered.* Cambridge, MA: MIT press.

Haig, D. (2007, December). Weismann Rules! OK? Epigenetics and the Lamarckian temptation. *Biology & Philosophy*, *22*(3), 415–428.

Heyes, C. (2012, August). New thinking: the evolution of human cognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *367*(1599), 2091–6.

Hinton, G., & Nowlan, S. (1987). How learning can guide evolution. *Complex systems*, *1*, 495–502.

Larson, E. J. (2004). *Evolution: The Remarkable History of a Scientific Theory*. New York: Modern Library.

Lloyd Morgan, C. (1896). On Modification and Descent. *Science*, *4*(99), 733–739.

Lumsden, C. J., & Wilson, E. O. (1981). *Genes, Mind, and Culture: The Coevolutionary Process*. Harvard University Press.

Mayr, E. (1963). *Animal Species and Evolution.* Cambridge, MA: Harvard University Press.

Odling-Smee, J., Laland, K. N., & Feldman, M. W. (2003). *Niche Construction: The neglected process in evolution*. Princeton University Press.

Osborne, H. F. (1896). A mode of evolution requiring neither natural selection nor the inheritance of acquired characteristics. *Transactions of the New York Academy of Science*, *15*, 141–148.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In S. Kirby & M. H. Christiansen (Eds.), *Language evolution: States of the art.* New York: Oxford University Press.

Pinker, S. (2010, May). The cognitive niche: coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences of the United States of America*, *107 Suppl*, 8993–9.

Pinker, S., & Bloom, P. (1990). natural language and natural selection. *Behavioral and brain sciences*, *13*, 707–784.

Reali, F., & Christiansen, M. H. (2009). Sequential learning and the interaction between biological and linguistic adaptation in language evolution. *Interaction Studies*, *10*(1), 5–30.

Richerson, P. J., & Boyd, R. (2005). *Not By Genes Alone: How Culture Transformed Human Evolution*. University Of Chicago Press.

Schuster, S., Wöhl, S., Griebsch, M., & Klostermeier, I. (2006, February). Animal cognition: how archer fish learn to down rapidly moving targets. *Current biology : CB*, *16*(4), 378–83.

Simpson, G. G. (1953). The Baldwin Effect. *Evolution*, *7*(2), 110–117.

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.

Whiten, A., & Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society of London B*, *367*, 2119–2129.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.