No evidence that age affects different bilingual learner groups differently: Rebuttal to van der Slik, Schepens, Bongaerts, and van Hout (2021)

Joshua K. Hartshorne
Department of Communication Sciences & Disorders
MGH Institute of Health Professions, USA

Abstract: Hartshorne, Tenenbaum, and Pinker (2018, A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 117, 263-277) presented the first direct estimate of how the ability to learn the morphosyntax of a second language changes with age, showing a sharp decline in late adolescence. Recently, van der Slik, Schepens, Bongaerts, and van Hout (2021, Critical period claim revisited: Reanalysis of Hartshorne, Tenenbaum, and Pinker (2018) suggests steady decline and learner-type differences. *Language Learning*) purport to show that in fact Hartshorne et al's (2018) data are better explained by a gradual decline in learning with age, at least for some types of learners. However, these conclusions are based a misunderstanding of their own analyses, which in fact do not test whether the decline in learning is sharp or gradual but whether it is asymmetric, slowing with time. After correcting conceptual and mathematical errors in their analyses, the results strongly confirm the original conclusions of Hartshorne and colleagues: every type of bilingual investigated shows a sharp drop in learning rate in late adolescence.

Keywords: critical periods; sensitive periods; L2

Corresponding author: Joshua K. Hartshorne, Department of Communication Sciences & Disorders, MGH Institute of Health Professions, Boston, MA 02129, USA. Email: joshua.hartshorne@hey.com.

ORCID ID: https://orcid.org/0000-0003-1240-3598

Citation: Hartshorne, Joshua K. (2024). No evidence that age affects different bilingual learner groups differently: Rebuttal to van der Slik, Schepens, Bongaerts, and van Hout (2021). *Language Development Research, 4*(1), 431–454. http://doi.org/10.34842/y8sz-4q83

Introduction

One hardly needs to conduct an experiment to show that people who begin learning a second language (L2) as an adult rarely if ever reach the same level of proficiency as those who start in early childhood, though plenty of experimental data do exist (Birdsong, 2018; Flege, 2019; Hartshorne, 2020b). What remains highly controversial is *why*: is poor achievement by later learners due to differences in neural plasticity or motivation, differences in the input, interference from a first language, or something else?

One way of constraining these possibilities is to measure the age at which learning rate starts to decline: if learning success begins to decline at age X, it is probably due to something that happened at age X, not at age Y.[1] Unfortunately, until recently there were no effective estimates of how learning rate changes with age. Measuring learning in the laboratory proved fruitless: During the initial stages of learning, older learners actually learn second languages faster (Asher & Price, 1967; Chan & Hartshorne, in press; Ferman & Karni, 2010; Krashen, Long, & Scarcella, 1979; Snedeker, Geren, & Shafto, 2012; Snow & Hoefnagel-Höhle, 1978). Decades-long longitudinal studies would work better but have not been done. Another approach, developed in the 1960s, is to find the oldest age at which someone can start learning a language and still reach native-like proficiency (Asher & García, 1969; Johnson & Newport, 1989). This method is limited by its inherent ambiguity: finding that (for instance) people who started learning a language at age 8 do better than those who started at 10 does not tell you much about when that difference appeared: the latter group might start off more slowly, or they may start off just as fast but fall behind after 3 years. Or 5. Or 10. In fact, it can be shown that such data do not constrain theory much if at all (for a more thorough exposition, see Appendix B).

Hartshorne, Tenenbaum, and Pinker (2018) (henceforth HTP) addressed the limitations reviewed above by applying a novel analytic model to a massive dataset of English morphosyntactic knowledge of 669,498 native and non-native English speakers, including monolinguals, simultaneous bilinguals, and second-language learners who either learned in an English-speaking country ("immersion learners") or not in an English speaking country ("non-immersion learners"). Morphosyntactic knowledge of each subject was assessed by a comprehensive 132-item self-paced written grammaticaly judgment and usage test (a complete list of stimuli are available in HTP's supplementary materials). Critically, the model (described below) disentangles how learning ability changes with age from other factors, including ceiling effects and years of exposure. The results indicated that the rate at which learners acquire English mor-

[1] Arguably, the rapidity with which learning rate declines is also informative. However, such arguments rest primarily on intuition, and intuitions can vary. For instance, while it is generally argued that a rapid decline in learning rate is most consistent with biological causes and that a slower decline would suggest social causes, a reviewer of this article made the opposite argument.

phosyntax declines substantially at around 17-18 years old, followed by an increasingly gradual decline into old age (Fig. 3).

Recently, van der Slik, Schepens, Bongaerts, and Hout (2021) (henceforth, *SSBH*) have challenged this conclusion. They compare HTP's model to an alternative model that they assert models age-related declines as gradual, not sharp. They report that their model "had a better fit when applied separately to monolinguals, bilinguals, and early immersion learners. Only for nonimmersion learners and later immersion learners did [HTP's] model have a better fit" (p. 87). They conclude that HTP's "overall conclusion of one sharply defined critical age at 17.4 for all language learners is based on artificial results" (p. 87).

In this paper, I show that these conclusions are based on conceptual confusions and mathematical errors. Critically, SSBH's alterrnative model does not require age-related declines in learning to be gradual. In fact, in every one of SSBH's analyses, their model either finds qualitatively the same result as HTP's original model or fits less well. Ironically, it is only for the late immersion learners and non-immersion learners that it finds a decline more gradual than what is inferred by HTP's model, but those are the cases in which HTP's model fits substantially better. Below, I show this with graphs and explain mathematically why this would be the case.

Regardless, the difference in results across different learner groups is mostly artifactual. Both HTP's and SSBH's models are designed to predict variability in learning outcomes based on age at which one starts learning the language. When applied to groups where everyone started learning at the same age (e.g., monolinguals or simultaneous bilinguals), they have difficulty distinguishing effects of experience from effects of age, complicating any conclusions.

Below, I first lay out the logic of HTP's model in more detail than can be found in the original paper. I then use this foundation to explain how SSBH's model works and how it differs from HTP's. It emerges that the difference between the models is not in whether one is "continuous" or "discontinuous" (the terms used by SSBH) but rather in terms of symmetry (HTP's model can detet a decline that is initially rapid and then slows down; SSBH's cannot). In this context, I re-present and re-evaluate SSBH's results, finding they strongly support HTP's original conclusions. However, it is conceivable that this finding is an artifact in limitations of SSBH's method. Thus, I test SSBH's hypotheses using a more precise model and a larger dataset. If anything, the results only more strongly support HTP's conclusions and militate against those of SSBH. These analyses also address the question that SSBH inadvertently tested, showing that the decline is indeed asymmetric: initially sharp and then more gradual.

A close look at HTP's model

HTP start by analyzing learning curves: knowledge as a function of years spent learning. Intuitively, if native speakers are more successful at learning a language than are late-L2 learners, their learning curves should be different. The fact that native speakers ultimately learn more means that their learning curves reach a higher asymptote. It is possible that the learning curves are steeper as well (they learn faster). By comparing learning curves for learners who began at different ages, it is possible to mathematically infer how learning changes with age, type of exposure, etc.

We describe the mathematical inference process below. First, we note that one important innovation of HTP was to empirically measure these learning curves. That is, they had enough monolinguals who had been speaking English for different lengths of time to empirically plot the monolingual learning curves. The same was true for simultaneous bilinguals, immersion learners who started at ages 1-3, and so on (Fig. 3A).

They note that, particularly for native speakers and early-L2 learners, the learning curve follows a very clear exponential decay (Fig. 1). Exponential decay is extremely common in natural processes. In this case, what it means is that how much a language learner learns at any given time depends on how much is left to learn. Formally, they model grammatical knowledge g at time t given that one started learning at time t_e (time of exposure) as:

$$g(t) = 1 - e^{\int_{t_e}^{t} -r dt} \tag{1}$$

where r is the learning rate. Note that if r is constant, the integral reduces to $(t - t_e)r$. Note that dt indicates that we are integrating over t.

This formula fits monolingual data almost perfectly by assuming that in each year, monolinguals learn a constant 13% of what is left to learn (Fig. 1, left). That is, they learn 13% in the first year, 11% in the second year [(100% - 13%) * 13%], 10% in the third year [(100% - 13% - (100% - 13%)*13%) * 13%], and so on. This is an asymptotic process and never quite ends, though after a certain amount of time the changes become negligible.

However, the curve reaches very different asymptotes for different learners: highest for monolinguals (Fig. 1, left), somewhat lower for simultaneous bilinguals (Fig. 1, center), and even lower for some groups of L2 learners (e.g., Fig. 1, right). More generally, HTP show that while there are main effects of whether the learner is monolingual or bilingual and whether they learned in an immersion or non-immersion setting, the learning curves are otherwise indistinguishable regardless of the age at which learning began, up to about 10 years of age: not only are the asymptotes the same, but the curve is just as steep (see also Fig. 3). For learners who begin later, however, the later they started learning English, the shallower the decay rate and the lower the asymptote.

One might suspect this could be accounted for by assuming different types of learners learn at different rates. Perhaps monolinguals are simply faster learners than are late L2 learners. HTP considered a model where r was fit separately for monolinguals, simultaneous bilinguals, immersion L2 learners, and non-immersion L2 learners. Formally, HTP introduced a parameter E, which was set to 1 for monolinguals and allowed to vary between 0 and 1 for each of the other three groups:

$$g(t) = 1 - e^{\int_{t_e}^{t} -Erdt} \quad (2)$$

HTP dubbed this E the "Experience discount factor", reflecting the intuition different learner groups get different amounts of exposure to English. However, this is an interpretation: mathematically, it simply means the learning rate is different across groups.

HTP report that this is insufficient, something we reproduce here with slightly different math: forcing the learning curves for different learner groups to share the same asymptote results in very poor fits (Fig. 1, dashed red lines). In intuitive terms, it is not just that simultaneous bilinguals or later learners learn more slowly than monolinguals, but that they stop learning before reaching monolingual proficiency.

HTP showed that this could be well-explained by assuming the learning rate (formally, variable r in the exponential decay curve) decreases as the learner ages. Specifically, they modified the exponential decay model above to one where the learning rate starts at a relatively high value. Specifically, they assumed that r is initially constant through some "critical" age t_c, after which it declines according to a sigmoid (an s-shaped curve). Thus, learning rate r at time t is given by:

$$r(t) = \begin{cases} r_0 & t \leq t_c \\ r_0(1 - \frac{1}{1+e^{-\alpha*(t-t_c-\delta)}}) & t > t_c \end{cases} \quad (3)$$

where r_0 is the initial learning rate, t_c is the critical age, and α and δ are parameters governing the steepness of the sigmoid and the location of its decline, respectively.

Critically, assuming that the decline in learning rate is sigmoidal was not so much a theoretical assumption as a lack of one. Sigmoids can decline slowly and gradually or

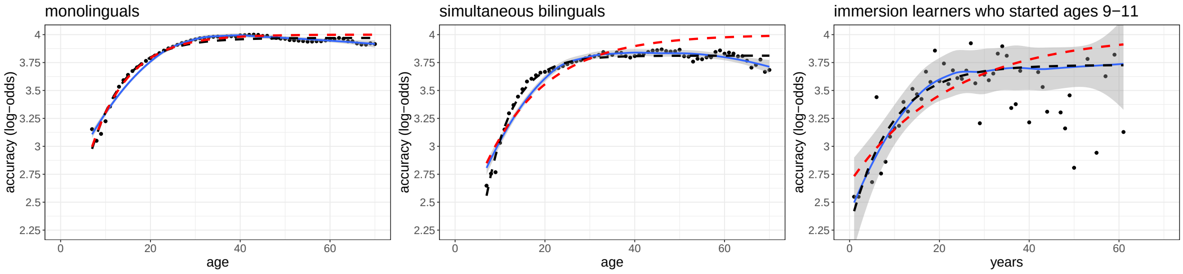

Figure 1. ***Y-axis shows accuracy on HTP's test using the empirical logit scale. Data is from HTP (Hartshorne, 2020a). Data for monolinguals (N=244,840; left), simultaneous bilinguals (N=30,347; center), and learners who began at ages 9-11 and learned in an immersion/immigration setting (N=1,373; right) are each shown. Performance is averaged by year (dots), with a LOESS curve in blue and a 95% confidence interval shown in gray. Results were fit to an exponential decay model with free parameters for intercept, asymptote, and rate (dashed black line) or with the parameter for rate fixed to be the same across all three learner groups (dashed red line). It is clear that the latter provides a poor fit.***

in a sharp step function (Fig. 2, Top) or not at all (simulating no age-related change). The decline can start at any point from birth to not at all.[2]

Sigmoids are mathematically convenient, having few parameters and being integrable. The latter is critical, since we must be able to perform integrals over r (see Equation (1)). Moreover, the parameters α and δ allow sigmoids to vary in how sharply they decline and where on the x-axis the decline occurs (Fig. 2, top row).

The is one crucial limitation to a sigmoid: It is symmetric. That is, if the decline begins quickly, it must reach floor quickly; if it reaches the floor slowly, the decline must begin slowly. By requiring r to initially be a constant (r_0) up until some age t_c, HTP circumvented this issue. This allows for declines that start rapidly but then level off (Fig. 2, second and third rows). Note that if $t_c = 0$, then the formula reduces to a standard, symmetric sigmoid. That is, HTP's model does not assume that learning declines quickly and then the decline levels off; it merely includes that as one of the hypotheses being tested.

Thus, by fitting the parameters to the data, HTP inferred the shape of age-related decline. As shown in Fig. 3, the model finds that the data were best explained by a sharp drop in learning rate at 17-18 years of age, followed by a more gradual decline. This fits the data quite well (compare Fig. 3 A&B with C&D). Not surprisingly, the model

[2]Strictly speaking, sigmoids are declining at every point along the x-axis, but for most of that span it is so mild as to be negligible. The exception is the degenerate case where $\alpha = 0$ and there is no decline at all.

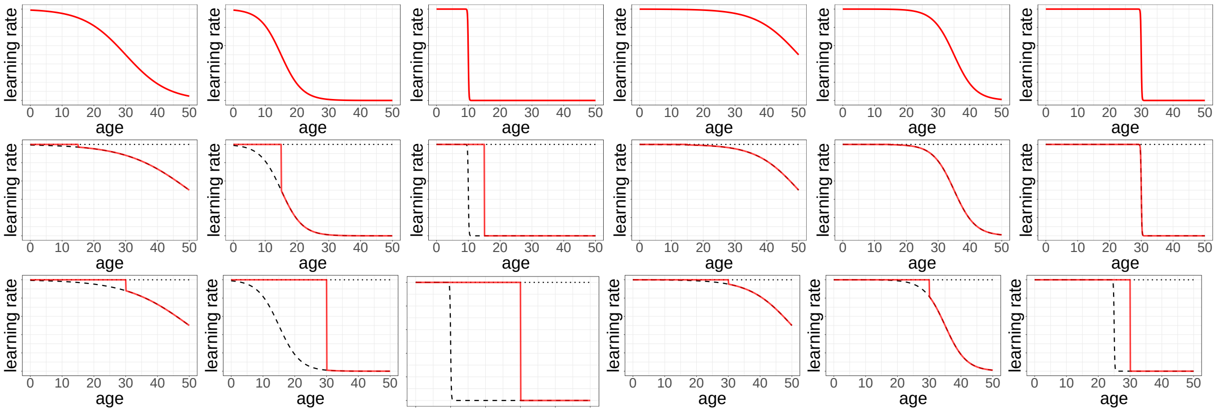

Figure 2. ***Top row: Sigmoids can decline more or less sharply (compare within the three left panels or within the three right panels) and can decline earlier or later (the three right panels are shifted versions of the three left panels). Second and third rows: HTP augmented sigmoids by composing them with a straight line, joined at age t_c. In these panels, the straight line is shown as a dotted line and the sigmoid is shown as a dashed line. The composed curved is outlined in red. Columns: Each column involves the same underlying line and sigmoid, but with different values of t_c, which is either 0 (top row), 15 (second row), or 30 (bottom row). Note that this figure was designed to illustrate the differences between HTP and SSBH and does not cover the full range of possible curves; for more examples, see HTP or Chen and Hartshorne (2021).***

fit much less well if learning rate (r) was forced to be constant across ages ($R^2 = 66\%$ vs. $R^2 = 89\%$; see Fig. 3, F, G, and H).

The fact that learning rate drops substantially with age explains two key empirical findings. Most importantly, it means that learning that happens in adulthood is quite slow, changing the effective asymptote. Although the model fit indicates that simultaneous bilinguals learn almost as fast as monolinguals (E is slightly less than 1), they are just a little behind. As a result, when the age-related decline kicks in at around 18 years old, they lose the ability to catch up. Later-learners are doubly-affected: their inferred learning rate (the product Er) is substantially lower than monolinguals', and they have fewer years of learning before speed slows and the asymptote lowers.

The second key empirical finding fit by the model is that learning curves learning curves continue to become progressively more shallow the later the learner began, even for learners who began learning after the critical age t_c (comp. Fig. 3 I & C). This cannot be explained if learning rate falls to a floor at t_c. Indeed, HTP report the results of such a model (a "step-function model"), and found that this fit less well, precisely because it fails to capture differences among learners who began learning at different intervals after t_c (Fig. 3 I, J, and K).

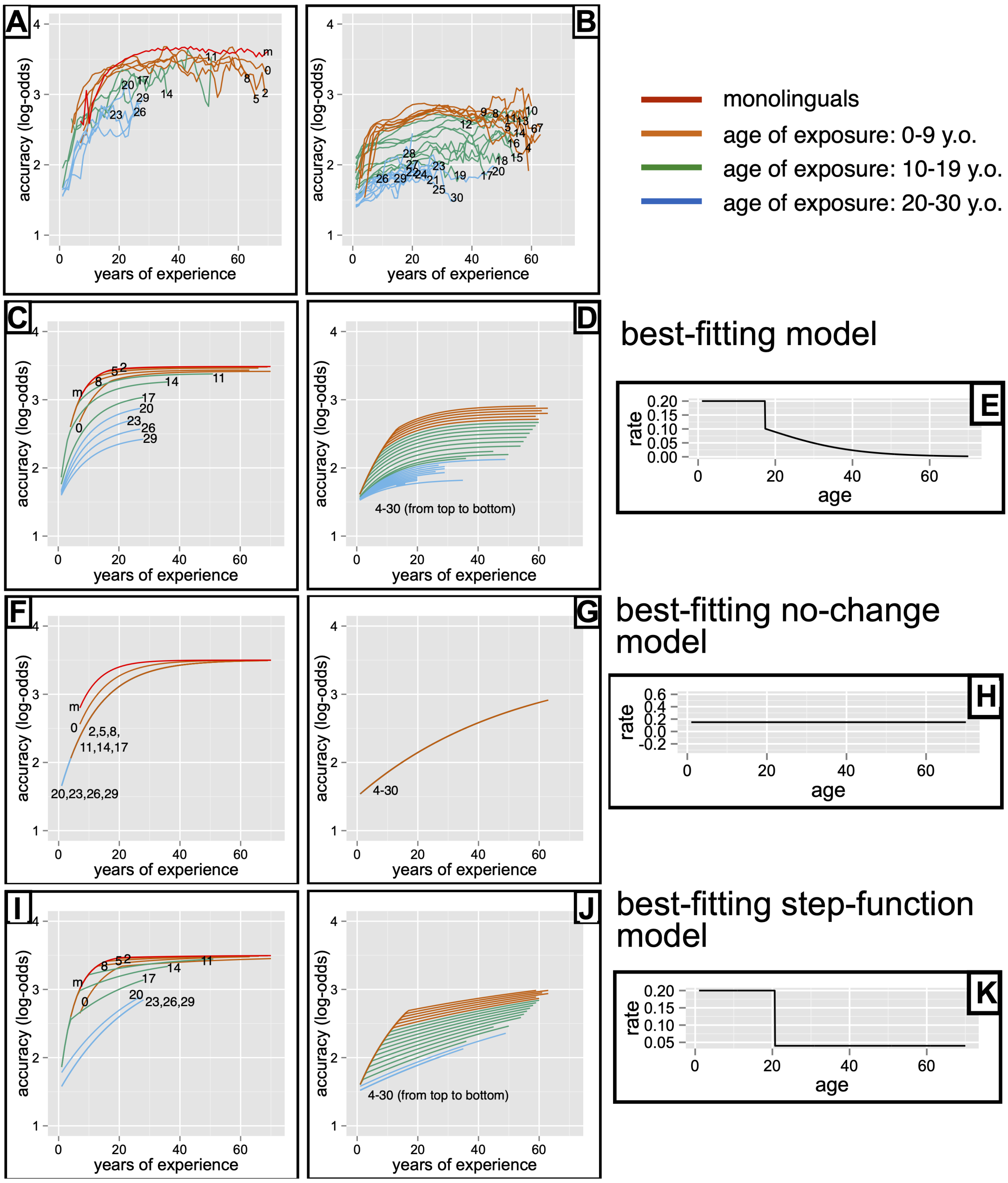

Figure 3. ***Figures from HTP, used with permission. Panels show the empirical results (top row), and the best fits for HTP's model (row 2nd from top) and two alternative models (bottom two rows). Monolinguals and immersion learners are plotted in the left panels (A, C, F, & I). Non-immersion learners are shown in the middle column (B, D, G, & J). In both the left and center columns, data/fits are plotted in terms of years of experience with the language, which makes the contrasts between models easier to see. Finally, panels in the right column show the models' estimated learning rate (r) as a function of age.***

Understanding SSBH's alternative analyses

As previewed in the introduction, van der Slik et al. (2021) (henceforth, *SSBH*) claim to show that HTP's findings are based on "artificial results" (p. 87). Specifically, they report that a sharply-defined drop in learning rate is true only for non-immersion learners and immersion learners who began learning at 10+ years of age ("late immersion learners"). They support this claim by comparing the fit of HTP's model to that of a "continuous" model independently for each of five learner groups: monolinguals, simultaneous bilinguals, early immersion learners (age of acquisition < 10), late immersion learners (age of acquisition >= 10), and non-immersion learners, finding that the continuous model fits the first three groups better while HTP's model fits the latter two better. They write, "The early immersion learners now share their continuous model with the monolingual and [simultaneous] bilingual learners. The later immersion learners share their discontinuous model with the nonimmersion learners, with a similar age boundary of 19.0 years" (p. 101).

In fact, their analyses do not support their claims and actually strongly support HTP's. There are a number of issues, but at the heart are a conceptual confusion and an empirical error.

The first conceptual confusion is that their "continuous" model does not assume gradual, continuous change in learning rate, nor does HTP's model assume a sharp discontinuity. As a result, which model fits better does not, by itself, speak to SSBH's question. (I return to what question it does speak to below.) Specifically, SSBH's "continuous" model is HTP's model where t_c is fixed at 0 (or sometimes fixed at 1; this varies across analyses for unexplained reasons; see Appendix C). By fixing t_c to 0, SSBH are forcing age-related decline to follow a sigmoidal shape.[3] Critically, as shown in the previous section, this does not constrain the decline to be early or late, or gradual or sharp. In fact, it does not even constrain there to be any decline at all (a degenerate sigmoid is flat horizontal). As a reminder, $t_c = 0$ in the first row of Fig. 2, and yet the third and sixth panels show fairly sharp declines in learning rate. Moreover, it is not the case that $t_c > 1$ means the decline is sharp. In the second row of Fig. 2 $t_c = 15$, yet the first and fourth panels show mild, gradual declines.

Thus, the model comparisons reported by SSBH do not address the question of when the age-related decline begins or how sharp it is, but rather whether the data are fit better if $t_c = 0$ or $t_c > 0$. Recall that HTP introduced the free parameter t_c to expand the range of possible age-related decline curves beyond strictly sigmoidal shapes. In particular, HTP wanted to include the possibility of a sharp decline followed by a more gradual decline — something that is impossible with a sigmoid only. Indeed, this is

[3] It is a little more complicated when $t_c = 1$, which could allow for a sharp drop in learning at the age of 1. However, one does not really require a study to know that this is not the case, and indeed it is not what any model finds.

exactly what HTP report when fitting their model to the full dataset: a sharp decline at 17.4 years old, followed by a more gradual decline. Note that SSBH criticize HTP's use of model-fitting rather than model-comparison, since the latter method takes into account number of parameters and HTP's model has an additional free parameter (t_c). However, in this case, they get the same result: HTP's model if preferred by several orders of magnitude, even correcting for the additional parameter (see SSBH Table 1 and surrounding text).

To recap: SSBH's analyses amount to asking whether including the t_c parameter provides a sufficiently better fit to justify the extra parameter. (It almost always fits the data better.) However, this does not by itself say anything about the sharpness of the decline. Making that determination requires looking at the actual inferred age-related change curves. Doing so paints a picture diametrically opposed to SSBH's conclusions (this is the empirical error).

SSBH do not include figures of the age-related change curves, so they are plotted them in Fig. 4. (Note that SSBH's paper contains a number of calculation errors. While these do not change the qualitative results, I use the corrected numbers throughout; see Appendix C for details.) While the full model (with t_c as a free parameter) provides at least as good a fit in all cases, this fit is not sufficiently better to justify the additional parameter for monolinguals (log-likelihood = 45.90 vs. 45.80; $AIC_{diff} = 1.90$; corrected relative log-likelihood: 2.44:1, favoring SSBH), simultaneous bilinguals (log-likelihood = 71.30 vs. 71.30; $AIC_{diff} = 2$; corrected relative log-likelihood: 2.73:1, favoring SSBH), and early immersion learners (log-likelihood = 112.92 vs. 112.86; $AIC_{diff} = 1.88$; corrected relative log-likelihood: 2.55:1, favoring SSBH).[4] However for simultaneous bilinguals and early immersion learners, this is a distinction without much of a difference: in both cases, there is a fairly sharp drop at around the same age regardless of whether one looks at HTP's free-t_c or SSBH's fixed-t_c model. Indeed, given that the two models fit the data roughly equally well, this is exactly what one should expect. SSBH's fixed-t_c model is very slightly smoother, but this is an artifact of the fact that during curve-fitting, SSBH forced the sigmoid sharpness parameter to be no greater than 1.0; had they relaxed that restriction, the decline would have been sharper.[5] In any case, the practical differences here are trivial (see right-hand side of Fig. 4)

The situation for monolinguals departs even further from SSBH's assertion of a "con-

[4]Akaike's Information Criterion [AIC; Akaike (1974)] is commonly used to compare models with different numbers of parameters. Formally, $AIC = 2k - 2 * log(likelihood)$, where k is the number of free parameters in the model. A reasonable rule of thumb is to choose the model with fewer parameters unless the more complex model improves the AIC by at least 4 (Burnham & Anderson, 1998).

[5]HTP's model achieves a sharper drop by use of the t_c parameter. Without a variable t_c parameter, the only way SSBH's model can achieve a sharp drop is through adjusting the sharpness of the sigmoid. This again illustrates that the relationship between the parameters and the theoretically-relevant curves is non-trivial.

tinuous" decline, in that SSBH's fixed-t_c model finds a sharp drop at around 50 years old, whereas HTP's free-t_c model finds no age-related change at all. This difference is an artifact of the fitting procedures used for SSBH: although in theory a sigmoid can be a straight line, SSBH restricted the parameter values such that this is unachievable (for details, see footnote).[6] This finding recapitulates what I showed in the previous section: monolingual learning curves are well-fit by exponential decay.

Does the finding with monolinguals nonetheless support SSBH's contention that there is no "sharply delimited critical period for normally developing monolinguals" (p. 102)? Not really. The models we are are discussing here try to estimate changes in learning rate due to *age* deconfounded from changes due to *years of experience*. Recall that we expect the total amount learned each year to decline as there is less and less left to learn (this is formally implemented as exponential decay). The problem is that since monolinguals all started learning English at the same age (0), age and amount of experience are full confounded and cannot be disentangled. By analogy, suppose a researcher wanted to investigate how children's height changes with age, and so measured the heights of a large number of children on their 5th birthdays. The researcher would find that age was completely unrelated to height, but only because the dataset was constructed that way. The models gamely try to disentangle age and experience anyway, but they have little to go on.

In principle, though, a decline in learning rate should still be detectable as a deviation from perfect exponential decay. In practice, this is very difficult. As shown in Fig. 1, by the mid-20s, both monolinguals simultaneous bilinguals are very close to ceiling. Thus whether or not there is a learning-rate decline in late adolescence will have at best subtle effects (see Fig. A1), making it difficult to detect muich less time exactly. We return to this issue below and show that more precise analyses with a larger dataset in fact suggest a decline starting in late adolescence, same as for other bilingual groups.

As alluded to above, HTP's free-t_c model fits better than SSBH's fixed-t_c model, even after correcting for number of parameters, for both the late immersion learners (log-likelihood = 127.53 vs. 72.39; AIC_{diff} = 108.29; corrected relative log-likelihood: $10^{23.52}$:1, favoring HTP's free-t_c model) and for non-immersion learners (log-likelihood

[6] As described in Fig. 2, the sigmoid shape parameters allow the decline to move left or right. This is governed by the variable δ in Eq. (3), which is the midpoint in the decline (the sharpness of the decline — and, hence, its effective starting point — is governed by α). Critically, the 0 point is t_c. Thus, if $t_c = 40$ and $\delta = 50$ the midpoint of the decline is at 90. In order to speed up computation, HTP limited δ to run from -50 to 50 and t_c to run from 0 to 40. This means that the decline could happen anywhere from the age of -50 to 90. Declines starting after the age of 70 do not affect our analysis, since we exclude subjects over the age of 70. (The reason we allow negative numbers for δ is that if t_c is 40, values for δ between -50 and 0 are meaningful; numbers below -50 are not. This can be easily demonstrated by experimenting with the function $plotr$ included in the reproducible manuscript.) When SSBH set t_c to 1, this restricted the ages in which the decline could happen from around -50 to around 50. Ideally, they would have expanded the available range for δ.

= 526.50 vs. 262.70; AIC_{diff} = -261.80; corrected relative log-likelihood: $10^{114.14}$:1, favoring HTP's free-t_c model) (see Fig. 4).

In summary, for every one of the five learner groups SSBH considers, their own analyses point to a sharply defined critical period. In no case do they provide statistical evidence for a "continuous" decline.

Their results do differ from HTP's in one important way, which is that the timing of the decline varies across the learner groups, with the critical period appearing earliest for late immersion and non-immersion learners, later for early immersion learners, even later for simultaneous bilinguals, and latest for monolinguals (though, as we explain above, this last fact is likely an artifact of how they fit their model).

The significance of this result is unclear. As just explained, age and experience are confounded for the latter three groups, making it difficult to accurately assess the effect of age. Nonetheless, with greater power and precision, we might be able to detect deviations from a constant learning rate even in these groups. I turn to this possibility in the next section.

Differences between Learner Groups, Redux

In order to more precisely compare age-related changes in learning rate within each learner groups, I made several improvements on SSBH's analyses. First, I used the more flexible model from Chen and Hartshorne (2021). Chen and Hartshorne (2021) provided an alternative formulation of HTP's model that allows for a wider range of age-related changes curves. While they found that this did not qualitatively change HTP's results, the model is more precise.

Second, I used the larger dataset reported by Chen and Hartshorne (2021), which has nearly half a million additional subjects, resulting in 319,565 monolinguals, 41,534 simultaneous bilinguals, 21,174 immersion learners, and 543,407 non-immersion learners.

Finally, inspired by Frank (2018), I reexamined how HTP and SSBH had defined the asymptote for exponential decay (that is, the asymptote that would be reached if learning rate remained steady). Prior analyses had assumed fixed the asymptote to 3.5, which HTP had determined by visual inspection. With the larger dataset, it was clear that this was a little too low, and that actual maximum hit by subjects is 3.66. (Frank (2018) suggests finding an analytic method that does not require specifying an asymptote. This seems like a very good idea but so far nobody has found one.)

I then fit the expanded dataset twice, either requiring the age-related changes in learn-

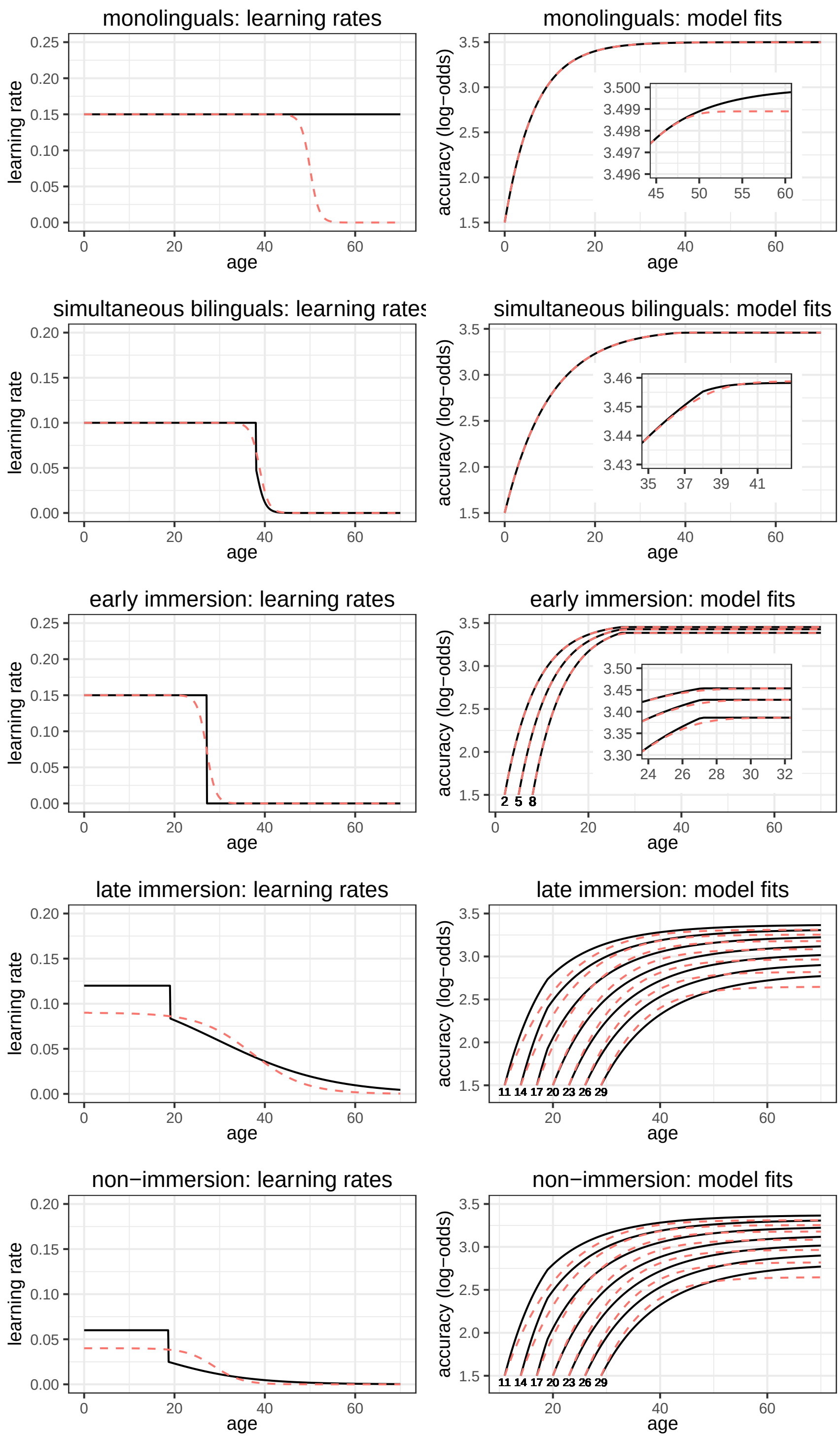

Figure 4. ***Model fits for each learner group, conducted separately.*** **Left panels*:*** ***Inferred age-related changes in learning curves.*** **Right panels*:*** ***fitted models (insets shows magnified view, where necessary).*** **Solid black*:*** ***HTP's model (t_c as a free parameter).*** **Dashed red*:*** ***SSBH's continuous model ($t_c = 1$).***

ing rate to be the same for all subject groups (Fig. 5, solid black lines) or allowing them to vary across the five learner groups defined by SSBH (Fig. 5, dashed red lines). Allowing for independent age-related change for each learner group actually led to a significantly *worse* fit after correction for number of parameters (AIC_{diff} = -14.49; relative log-likelihood: 1402 to 1). In any case, fitting each group separately nontheless reliably results in a sharp decline in late adolescence, with the sole exception of monolinguals, where the decline was much later Fig. 5. However, given that monolinguals are quite close to ceiling by late adolescence, the uncertainty about exactly where the decline is likely to be substantial, so one should be cautioned against making too much of this result even if it were significant, which it is not.

In short, there seems little reason to suppose more than minimal differences in age-related change in learning rate across the different learner groups, at least in this dataset.

Summary and Conclusions

SSBH suggest that different learner groups exhibit distinct effects of age on learning, with some groups declining continuously from birth (monolinguals, simultaneous bilinguals, early immersion learners) while others show a marked decline in adolescence (late immersion learners, non-immersion learners). However, none of their arguments or results hold up under scrutiny. If anything their results support HTP's conclusions more than their own. A revised analysis, more appropriate to SSBH's questions, supported HTP's conclusions even more strongly.

This does not mean that the case is closed. Although the dataset and the analyses support a rapid drop in morphosyntax learning ability in late adolescence, there are significant limitations. While a lot of care went into creating HTP's quiz, I doubt that any 95-question quiz can assess an infinitely expressive human grammar precisely and without bias. If such a quiz can be created, we certainly lack the theoretical understanding of morphosyntax needed to construct it at the moment. Moreover, HTP's quiz probes meta-linguistic grammaticality judgments. This is certainly an important linguistic phenomenon – the spectacular failure of late learners to acquire native-like meta-linguistic knowledge is part of what we wish to explain! – but it clearly involves cognitive mechanisms not required for other linguistic phenomena, which themselves depend on cognitive mechanisms not needed for meta-linguistic judgment. To the extent this mechanisms are themselves affected by age, the picture will depend on which phenomenon we study.

In terms of the analytic model, even the best of the models explored above do not fit the data perfectly, and they alide some known issues. It has no previsions for senescence, which turns out to begin much earlier than had been visible in HTP's data (compare Fig. 3A with Fig. 5, top), starting perhaps as early as the mid 40s. Unfortunately, our original

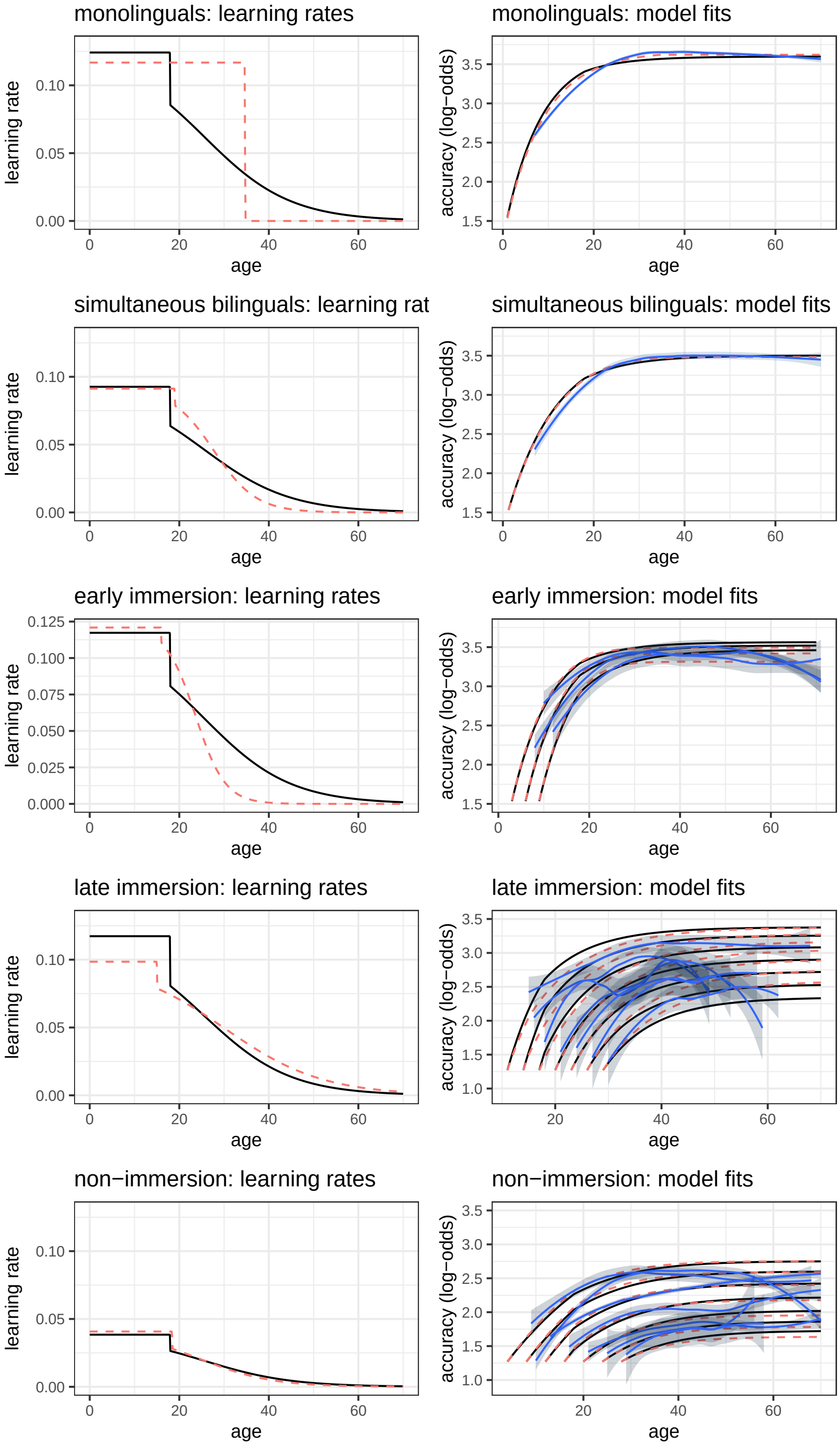

Figure 5. ***Comparisons of the revised model trained on all data (solid black lines) and on individual learner groups (dashed red lines), and LOESS-smoothed data (blue lines with shaded 95% confidence intervals). For nonimmersion learners, only a subset of curves are shown.***

strategy of simply excluding older subjects will not work: excluding subjects older than 45 would vitiate our ability to study later learners. Similarly, the model cannot entertain age-related *increases* in learning rate during childhood, even though these are clear in the present data and in prior work (Asher & Price, 1967; Chan & Hartshorne, in press; Ferman & Karni, 2010; Krashen et al., 1979; Snedeker et al., 2012; Snow & Hoefnagel-Höhle, 1978). Relatedly, the models assume that age-related change is driven by a single underlying factor. This is certainly at least somewhat wrong; in the limit, the r curves we see could be an epiphenomenon of age-related changes in many different underlying mechanisms, each of which looks quite different. Finally, the models assume learning is asymtotic, whereas Frank (2018) correctly notes that many modern theories (especially construction grammars) posit that the set of grammatical structures is a) unbounded, and b) a moving target due to language change.

More broadly, as highlighted by HTP, our analyses estimate age-related change in learning rate. They cannot speak to whether this represents biologically-determined change, age-related changes in environment, or something else. The results certainly constrain the possibilities (e.g., factors that do not change rapidly in late adolescence are unlikely to explain the results), but ultimately that evidence is indirect. We need studies directly testing the causal role of candidate influences on learning.

All of which is to say that HTP and follow-up papers (Chen & Hartshorne, 2021; Frank, 2018; Hernandez, Bodet, Gehm, & Shen, 2021; van der Slik et al., 2021) are just the start of a conversation. We will need many more studies of similar scale and scope to resolve the open theoretical questions.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Asher, J. J., & García, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal, 53*(5), 334–341.

Asher, J. J., & Price, B. S. (1967). The learning strategy of the total physical response: Some age differences. *Child Development*, 1219–1227.

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology, 9*, 81.

Burnham, K. P., & Anderson, D. R. (1998). Practical use of the information-theoretic approach. In *Model selection and inference* (pp. 75–117). Springer.

Chan, J., & Hartshorne, J. K. (in press). *Is it easier for children to learn english if their native language is similar to english?* Cascadilla Press.

Chen, T., & Hartshorne, J. K. (2021). More evidence from over 1.1 million subjects that the critical period for syntax closes in late adolescence. *Cognition, 214*, 104706.

Ferman, S., & Karni, A. (2010). No childhood advantage in the acquisition of skill in using an artificial language rule. *PloS One, 5*(10), e13648.

Flege, J. E. (2019). A non-critical period for second-language learning. *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn, Aarhus University. Open Access e-Book at Aurhus University Library*.

Frank, M. C. (2018). With great data comes great (theoretical) opportunity. *Trends in Cognitive Sciences*, *22*(8), 669–671.

Hartshorne, J. K. (2020a). *Data: A critical period for second language acquisition: Evidence from 2/3 million english speakers*. OSF. Retrieved from osf.io/pyb8s

Hartshorne, J. K. (2020b). How massive online experiments (MOEs) can illuminate critical and sensitive periods in development. *Current Opinion in Behavioral Sciences*, *36*, 135–143.

Hartshorne, Joshua K. and Tenenbaum, Joshua B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Hernandez, A. E., Bodet, J. P., Gehm, K., & Shen, S. (2021). What does a critical period for second language acquisition mean?: Reflections on Hartshorne et al. (2018). *Cognition*, *206*, 104478. https://doi.org/10.1016/j.cognition.2020.104478

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of english as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, 573–582.

Snedeker, J., Geren, J., & Shafto, C. L. (2012). Disentangling the effects of cognitive development and linguistic expertise: A longitudinal study of the acquisition of english in internationally-adopted children. *Cognitive Psychology*, *65*(1), 39–76.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 1114–1128.

van der Slik, F., Schepens, J., Bongaerts, T., & Hout, R. van. (2021). Critical period claim revisited: Reanalysis of hartshorne, tenenbaum, and pinker (2018) suggests steady decline and learner-type differences. *Language Learning*.

Data, Code and Materials Availability Statement

Data and analysis code can be found at osf.io/u6fq5.

Ethics statement

This paper does not contain human subjects research.

Authorship and Contributorship Statement

JKH is solely responsible for this work.

Acknowledgements

The author thanks F. W. P. van der Slik, Tianhua Chen, David Birdsong, Lisa Pearl, Michael C. Frank, Emma Marsden, two anonymous reviewers, and the attendees of talks at the Reading/Tromso joint bilingualism meeting, the BMB Lab at UC-Irvine, the Stanford Language and Cognition Lab, the Brain and Language Lab at Georgetown University, and the Institute of Linguistics and National Taiwan University for comments and suggestions. The author declares no competing interests.

Appendices

Appendix A: Supplementary Figures

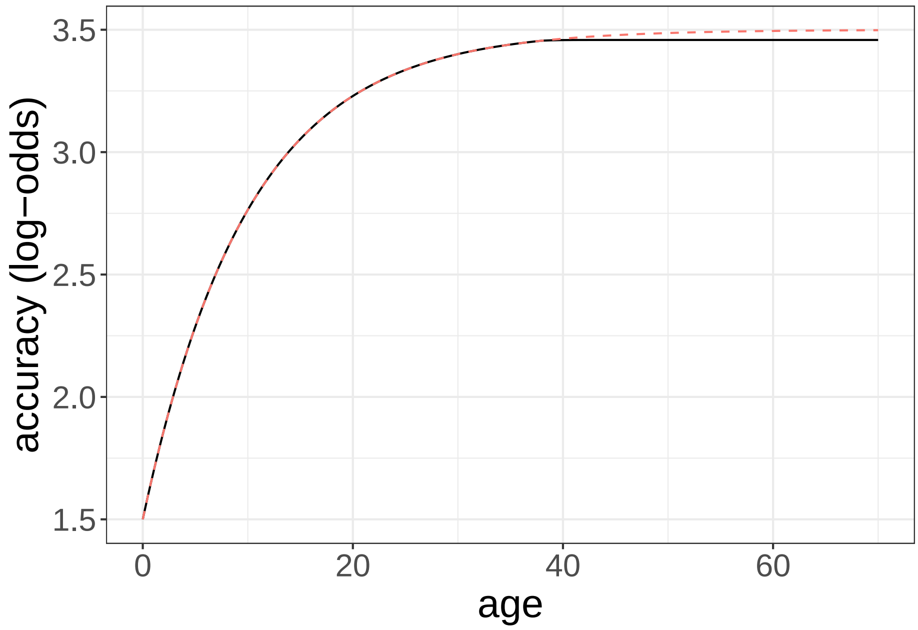

Figure A1. ***When fit to the simultaneous bilinguals only, both the SSBH and HTP models suggest a decline in learning rate at around the age of 40. Here, we plot the predicted learning curves for with (solid black) and without (dashed red) that age-related change. As can be seen, the differences are quite small.***

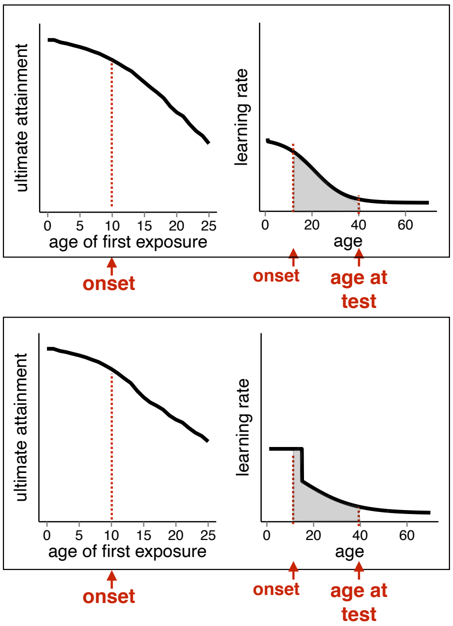

Figure A2. ***Each point on an ultimate attainment curve (left panels) is related to an integral under the learning rate curve from onset of learning through the age at test (shaded portion of right panels). Most studies have assumed that it is possible to infer the shape of the theoretically-critical learning rate curves from easier-to-measure ultimate attainment curves. However, simulations from HTP show that highly similar ultimate attainment curves (left: top vs. bottom) can actually be explained by very different learning rate curves (right: top vs. bottom).***

Appendix B: Limitations of Studying Ultimate Attainment

The oldest age at which one can start learning a new language and still reach nativelike proficiency is not (necessarily) the age at which learning rate declines. To give an intuitive example, suppose we know that if Agnes leaves her home at 8:15 in the morning, she makes it to work comfortably before 9:00. If she leaves after 8:15, she runs into a traffic jam and arrives much later. Does this mean that the traffic picks up at exactly 8:15? Perhaps. Even a slight decrease in speed, applied over the entire travel distance, could be enough to make her tardy. Alternatively, the traffic may grind to a halt at 8:45, so if Agnes hasn't arrived by then, she is out of luck. The point is that if we know what time she left home and how far she got, we know her *average* speed, but not her speed at any given point along the way.

Similarly, if Bartholomew starts learning Swahili as an adult and manages only 80% the proficiency of a native speaker, this does not mean that he started out learning more slowly than a Swahili-acquiring infant. In fact, as mentioned above, during the initial stages of learning, older learners actually learn second languages faster (Asher & Price, 1967; Chan & Hartshorne, in press; Ferman & Karni, 2010; Krashen et al., 1979; Snedeker et al., 2012; Snow & Hoefnagel-Höhle, 1978). Thus, all we know for sure is that at some point along the way, his learning rate decayed to the point where he ultimately was unable to get to the finish line.

More formally, very different age-related changes in the ability to learn language can give rise to indistinguishable ultimate attainment curves (Fig. A2).

Appendix C: Additional errors and imprecisions in SSBH

SSBH make a number of factual misstatements and mathematical errors. The following list may not be exhaustive.

SSBH use Akaike Information Criterion (AIC) for model comparison, but in almost every case appear to have miscounted the number of parameters in the models (a key part of calculating AIC). For most of their analyses, the "continuous" model has 4 free parameters (r_0, α, δ, and the error variance), though in all but one case, they count it as having 5. The "discontinuous" model has one additional free parameter (t_c) but for some reason is counted as having 7. The exceptions are as follows: In the case of the monolingual analysis, they correctly assign the continuous model 4 parameters, but again over-count the discontinuous model (6 instead of 5). When fit to all data, there are 3 additional parameters (the three E parameters), which should give the "continuous" model 7 parameters (which they code correctly) and give the "discontinuous" model 8 (they count 9).

(Note that they explain in Footnote 5 that "the discontinuous model needs to fit three components, the continuous model only one (cf. Figure 1). That explains the difference of two degrees of freedom." It is not possible to count degrees of freedom by inspecting a graph, and the numbers here do not match the numbers in their code.)

These errors tend to overstate the evidence for the "continuous" model. For instance, the relative likelihood for the monolingual analyses in their Table 3 is reported as 0.16. Using the correct number of parameters, it is 0.41. That is, using AIC correctly, rather than the "continuous" model being nearly 7 times more likely, it is only about 3 times more likely. (Strangely, using SSBH's counting of parameters, the ratio is actually 0.15; I have not yet identified the source of that error.)

As described in the main text, the "continuous" model is simply the HTP model (which they call "discontinuous") with the t_c parameter fixed. Across analyses, it is sometimes fixed to 1 and sometimes to 0. SSBH do not provide any explanation, and indeed do not even mention this variation. Inspection suggests that the choice of 1 or 0 probably does not make much difference, though I did not test this systematically. Note that strictly speaking SSBH's "continuous" model is only a special case of HTP's model when t_c is set to 1, since HTP fit HTP's model with a restriction that $t_c > 0$.

SSBH report that HTP defined immersion learners as either simultaneous bilinguals or "later learners who spent at least 90% of their life in an English-speaking country" (SSBH, p. 7). In fact, later immersion learners were required to have spent at least 90% of their life *since starting to learn English* in an English speaking country (HTP, p. 266). This makes a considerable difference: analyses include immersion learners who began learning English as late as 30, so under SSBH's definition they would need to be at least 300 years old at time of testing. Similarly, SSBH incorrectly report that non-immersion learners were those "who spent at most 10% of their life in an English-speaking country" (SSBH, p. 8), whereas the actual definition is "spent at most 10% of post-exposure life in an English-speaking country and no more than 1 year in total" (HTP, p. 266). Note that SSBH do use the correct definitions in their own analyses, so this does not affect their results.

Probably because of their confusion about how subject groups were defined, SSBH mistakenly report that "more than 100,000 language learners in the HTP database could not be classified as belonging to one of the four groups *because key information was missing*" (emphasis added; p. 20). They assert that this high rate of missing data should cast doubt on the validity/accuracy of the HTP data. However, these subjects were not excluded for missing data but rather for having amounts of immersion intermediate between the "immersion" and "non-immersion" learners (see sentence spanning pp. 266-267).

SSBH misdescribe the stimuli. They report that HTP's test included 132 items, of which 95 were used for analysis "based on the criterion that at least 70% of the native English-speaking adults gave the same response" (SSBH, p. 8). In fact, the criterion was that the same response was given by at least 70% of native English-speaking adults in each of 13 dialect groups (HTP, p. 267). The reason was to exclude items for which there was significant dialectal variation. They also assert that HTP measures accuracy on the grammaticality judgment test on a scale of 0 to 1, reflecting "a proportion of correct answers (g)" (SSBH, p. 7). In fact, *g* represents log-odds accuracy on HTP's syntax test and runs from 1.5 to 3.5 (see HTP Supplementary Materials, p. 2). They misstate how HTP (and, it appears, they) calculated log-odds, asserting that it was based on proportion (*log(p/[1-p])*) (SSBH, p. 7) rather than the empirical logit transformation (*log((num correct+.5)/(num incorrect+.5))*).

In Table 2 and surrounding text, they report some discrepancies between the number of subjects per condition for the critical analyses reported by HTP (p. 266) and in SSBH's own analyses. The problem seems to be that they ran their exclusions in a different order from HTP. Specifically, both papers bin subjects by age, age of acquisition, and condition. We then restrict analyses to consecutive ages for which there were at least 10 participants in a 5-year window. HTP excludes subjects over the age of 70 before this binning, whereas SSBH exclude subjects over the age of 70 *after* binning. This means subjects over the age of 70 count towards binning for SSBH but not for HTP, allowing inclusion of more bins for SSBH. Thus, as they report, they end up with 38 more total included subjects. Since we provided them with the original code, it is not clear why they were unaware of these differences.

When replicating one of HTP's analyses, they report that they obtained "a slightly higher R^2 value of .92 (HTP found .89)" (SSBH, p. 10). This likely reflects the fact that while HTP report cross-validated R^2 values in order to address over-fitting, SSBH do not. This will necessarily result in higher R^2 values. In a personal communication, van der Slik suggested that because they ran the optimization algorithm for more iterations than HTP did, this should obviate the need for cross-validation. This is exactly backwards. It is a necessary fact that the more closely the model is fit to the data, the worse over-fitting gets. In any case, the result is that their R^2 values must be treated with caution: a particular model may achieve a better R^2 simply due to overfitting.

In Footnote 7, they write that Chen and Hartshorne "did not test if the application of their segmented model has resulted in a significant improvement in model fit as compared to the continuous model or even the original HTP discontinuous model." In fact, we provided two such metrics. First, the model fits available to ELSD are a proper subset of those available to Chen & Hartshorne's segmented sigmoid model, and thus fitting the revised model is *per se* a comparison of model fit. Second, Chen and Hartshorne also provide cross-validated R^2 statistics for both their model and the HTP model, allowing direct comparison.

While SSBH present these differences between early and late immersion learners as a novel observation, they were reported first by HTP. In particular, HTP in fact reported two sets of analyses showing that immersion learners who began before the age of 10 learn at least as rapidly and successfully as simultaneous bilinguals (HTP p. 270).

Note that I did not rerun SSBH's model fits themselves and instead copied those numbers from their tables. I cannot guarantee they are correct.

License