LINGUIST 697K: Introduction to Psycholinguistics

Lab Report #1 - Speech Perception

Christian Muxica

University of Massachusetts Amherst

Introduction

Psycholinguists describe language perception primarily as an information processing task- this generally a unifying perspective across the cognitive sciences. This task is conducted using a combination of both top-down and bottom-up information. Within the domain of language, bottom-up information constitutes the linguistic input which a perceiver receives. This most often comes in the form of the acoustic signals which define the spoken word. Top-down information on the other hand describes grammar- the set of abstract rules which the competent speaker of any given language is said to have acquired. A topic of perennial importance is the relative contributions of these two kinds of information to any given language act. One salient example being the literature of phonetic perception, both within and across languages.

In the act of comprehending spoken language, the listener must convert a continuous acoustic signal into discrete abstract linguistic representations. Different languages and dialects carve out this continuous acoustic space into discrete phonemes differently. This top-down grammatical information from a native language has been demonstrated to impact the bottom-up perceptions of speakers.

Pisoni & Tash (1974) record reaction times of English speakers in identification and discrimnation tasks using speech sounds along a /ba/ <--> /pa/ continuum of voice-onset time. In identification, Pisoni & Tash (1974) observe the speed and consistency of participant judgements decrease drastically as the VOT value of the sounds approaches (what appears to be) a perceptual threshold. At this value of VOT (approximately +20-30ms), participant judgements are inconsistent between categories and take around 80ms on average to respond. This is in contrast with average reaction times of around 25-35ms to respond outside this +20-30ms VOT range. Utilizing an A-X discrimination task, again along a /ba/ <--> /pa/ continuum, Pisoni & Tash (1974) find a similar perceptual threshold. When there is no difference or a large difference in VOT between the two stimulus, reaction times to label the sounds as the same or different respectively are (< 320ms) relatively fast. When there is a small difference in VOT (20ms) between the two stimulus, reaction times are (> 360ms) relatively slower.

Pisoni & Tash (1974) take these findings to support a three step model of speech perception in discrimnation experiments. In the first stage the perceived pair of stimuli are encoded in memory, in the second stage the raw acoustic similarity of the stimuli pair is evaluated, and in the third stage the phonetic similarity of the stimuli pair is evaluated. This derives the key finding from the discrimination experiments. The acoustic similarity or difference is detected early in the second stage for the acoustically identical or distant pairs of stimuli. The differences between the more similar stimuli pairs on the other hand require abstraction to phonemic categories in the third stage before discrimination can take place. With these findings, Pisoni & Tash (1974) demonstrate that the act of speech perception (or minimally discrimination) makes use of both top-down and bottom-up information.

Kazanina et al (2006) investigate a related cross-linguistic question. Both Russian and Korean speakers utilize the phones [d] and [t], but these sounds hold different status in each language. In Russian [d] and [t] encode a meaning contrast and thus correspond to two distinct phonemes, /d/ and /t/ in the mental grammar of speakers. This is not the case for Korean where [d] and [t] are simply allophones of one phoneme, /T/ which is realized as [d] intervocalically and [t] at word onset. Kazanina et al (2006) utilize a magnetoencephalography (MEG) oddball paradigm to investigate the impact of these grammatical differences on speech perception. In the oddball paradigm a series of sounds is dominated by one speech sound and sparsely filled with a different speech sound. If a difference is detected between the two speech sounds by a speaker, a mismatch negative (MMN) will be observed in the MEG data beginning at the onset of the sparsely played phone.

Kazanina et al (2006) perform this experiment on native Russian and Korean speakers with speech sounds along a [da] <--> [ta] continuum of voice onset time. What Kazanina et al (2006) observe is that despite the similar acoustic distribution of [da] and [ta] in both languages, only the neural activity of Russian speakers contained an MMN response. This indicates that the encoding of meaning which defines phonemes in a language has an almost immediate impact on preattentive speech perception. These findings pose a challenge for bottom-up theories which attempt to define perceptual categories purely in terms of the acoustic distributions of sounds speakers experience. Kazanina et al (2006) demonstrate that even preattentive speech perception is defined by top-down information such as meaning contrasts and phonemes.

The present research combines the approaches of these two papers. Kazanina et al (2006) investigated the effect of native language on preattentive speech perception utilizing languages which differ in perceptual categories. In this report, the effect of native language on speech perception is investigated utilizing English and Russian. These two languages both encode the speech sounds [da] and [ta] as separate phonemes, but the VOTs for these phonemes are quite different. The impact similar perceptual categories with different acoustics might have on the attentive measures utilized by Pisoni & Tash (1974) is unclear. In the present experiment, an English speaker is administered identification and discimnation tasks in both languages.

Method

I. Participants

One participant was utilized in the experiment. This participant was the principal investigator for this research as well. The participant was a 21 year old, male, undergraduate linguistics student enrolled at the University of Massachusetts Amherst. The participant was a monolingual Native English speaker from Massachusetts with no significant linguistic experience with the Russian language.

II. Apparatus & Materials

The only physical materials utilized were a single L450 ThinkPad laptop and a pair of Audio Technica ATH-M50x headphones. Four separate scripts were utilized for the discrimination and identification epxeriments for each language. All of these scripts were written in the python package PsychoPy 3 and run on the aforementioned hardware. These scripts were written originally by Collin Phillips and were taken directly from his personal website. The English speech sounds span from a VOT of 0-60ms and were each generated by Collin Phillips artificially with a speech synthesizer. The Russian speech sounds range in VOT from +10ms to -44ms, each was generated by a native Russian speaker and computer-edited to create the VOT continuum.

III. Procedure

All four experiments were completed in one 50 minute session by the single participant. In the identification experiments a randomized sequence of speech sounds varying in VOT was played through the headphones. The participant was instructed to identify each sound as belonging to either the /d/ or /t/ category. This judgement was provided using the ‘f’ and ‘j’ keys on the keyboard of the laptop. Reaction times to provide these judgements were recorded along with the response. The participant was given a maximum of six seconds to respond before any given trial would time out. In each identification experiment the participant listened to exactly 100 speech sounds. This process was identical for both languages.

In the discrimination experiments two speech sounds varying in VOT were played one after the other through the headphones in a randomized sequence. The participant was instructed to identify whether the two sounds played were the identical or different. This judgement was provided using the ‘f’ and ‘j’ keys on the keyboard of the laptop. Reaction times to provide these judgements were recorded along with the response. Again, the participant was given a maximum of six seconds to respond before any given trial would time out. In each discrimination experiment the participant listened to exactly 222 speech sounds. This process was identical for both languages.

First the identification experiments were run, followed by the dicrimination experiments. In each case the English experiment was run before Russian. In between each experiment a 5 minute break was provided to prevent participant fatigue.

Results

I. Identification Task

The results of the identification task for English speech sounds is largely similar to the findings observed by Pisoni & Tash (1974). As shown in Table 1, when the VOT is low from the range of 0-10ms the participant responds that the stimulus belongs to the /da/ category at ceiling. This high probability abruptly drops down close to chance once the VOT of the stimulus hits the perceptual threshold of 20ms. Any point beyond this threshold and the participant responds that the stimulus belongs to the /ta/ category with high probability. There are some /dae/ responses beyond the 20ms threshold, however these are marginal and never exceed 25% of responses. In Table 2 we can see reaction times for English identification. The lowest reaction times can generally be observed at the extremes of VOT. At 0ms and 10ms of VOT delay reaction times are the lowest- 792ms and 771ms respectively. This correlates with the 100% /da/ identification at these values as well. Reaction times peak close to the aforementioned perceptual threshold, between 24-28ms of VOT delay reaction times reach 1135ms and

1150ms on average respectively. Notably while reaction times decrease after the perceptual threshold

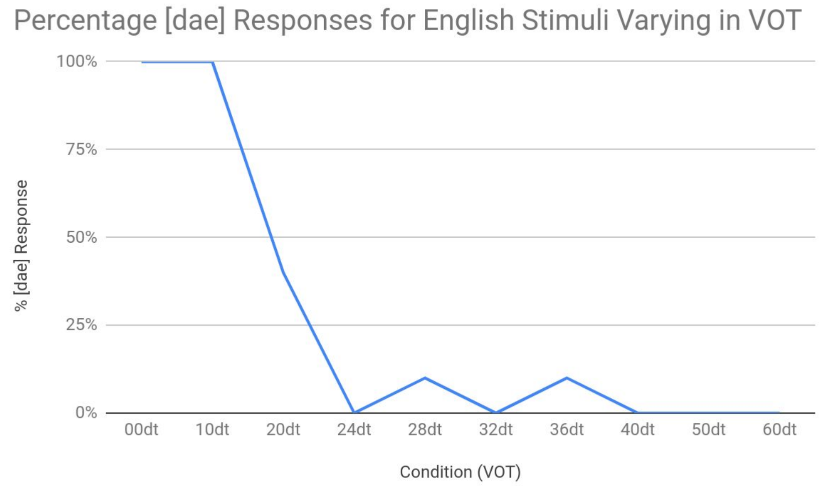

Table 1.

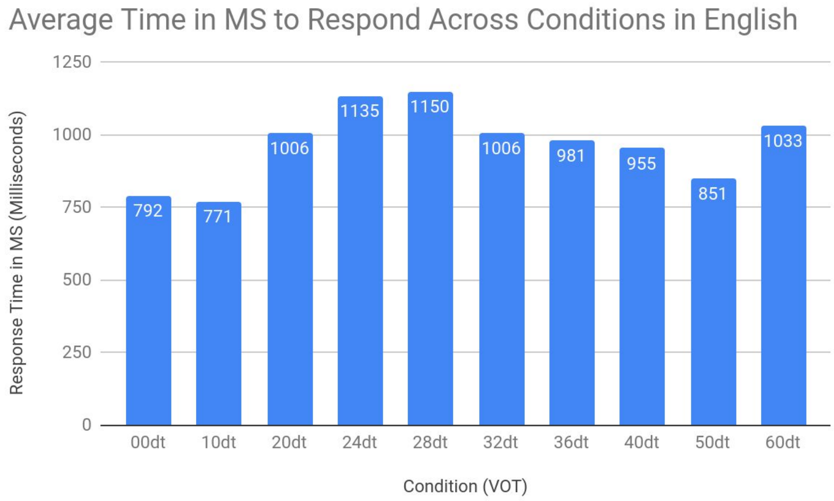

Table 2.

they never reach a low similar to that from the /da/ identified stimuli. In fact reaction times increase again at the 60ms of VOT counter to previous research.

The identification results for the Russian stimuli are less clear overall. Table 3 demonstrates that the two stimuli which lacked prevoicining yielded the highest rate of

/ta/ responses. Close to 75% of total responses for those conditions. Percentage of /da/ responses at -6ms VOT is about 50% or at chance for this paradigm. All prevoiced VOT values earlier than -6ms yield a high percentage of /da/ responses, never dropping below 75% for any stimulus. Reaction times seem to pattern with /da/ response percentage. Responses are fastest on average for stimuli which lack prevoicing. However, even these values are ~200ms slower than the fastest average response

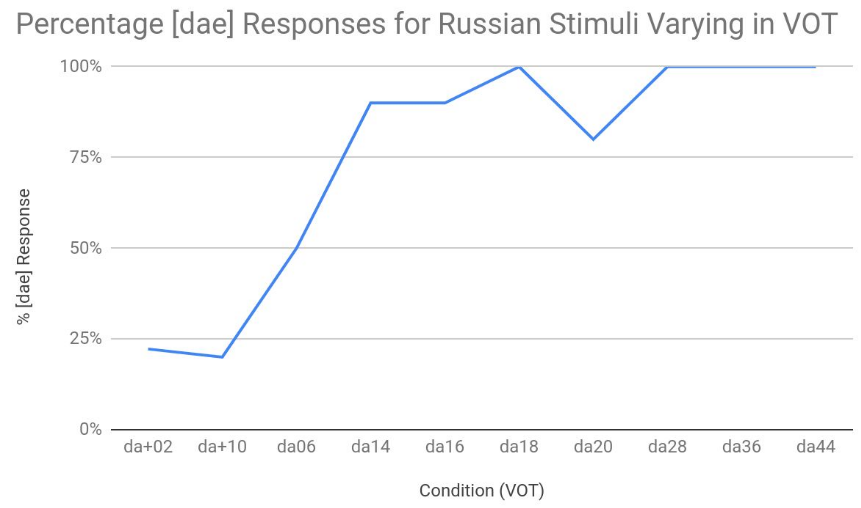

Table 3.

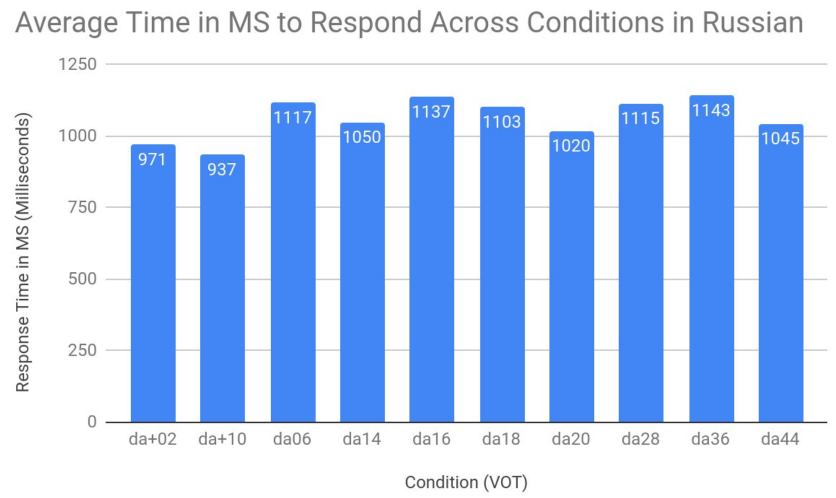

Table 4.

times for the English stimuli. There does not appear to be any discernible pattern for reaction times within those stimuli which are prevoiced, aside from that they are all higher than the prevoiced stimuli. Even at -6ms VOT, where identification was least certain, average reaction times are not notably different from any of the other prevoiced conditions.

II. Discrimination Task

Viewing the results of the English discrimination experiment by the step distance in VOT between the two stimuli presented yields a clear pattern in the discriminability index. Looking at Table 5, we see that identical stimuli with 0 step VOT distance have the highest discriminability index. This indicates that the signal received from this condition was strong and correspondingly the participant was accurate. This is in

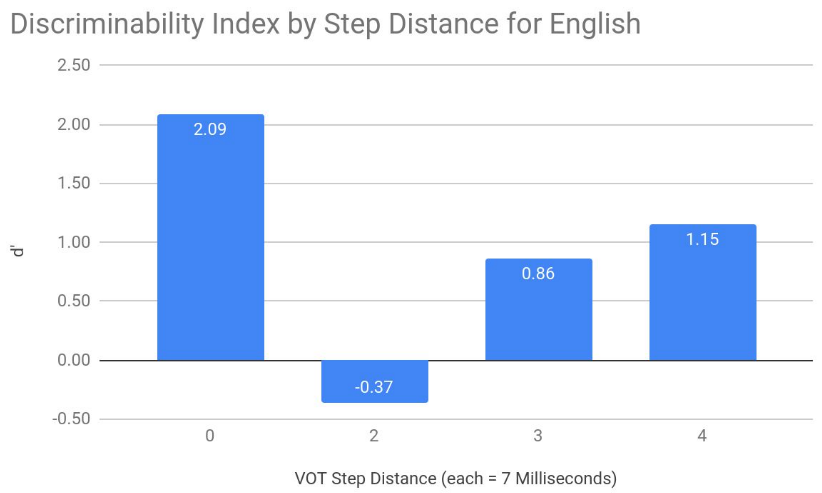

Table 5.

contrast with the poor signal from the not identical, but not acoustically distant 2 step stimuli. From there discriminability increases with the step distance between stimuli, but never reaches the height of the 0 step identical stimuli.

The average reaction times for each of the step distance conditions presents a similar pattern. As seen in Table 6, the acoustically close 2 step condition yields the the slowest reaction time- about 100ms longer than the other conditions. The reaction times are fastest and nearly identical for the 4 and 5 step distance stimuli. In contrast with aforementioned discriminability results, the identical 0 step stimuli reaction times were on average slower than the 4 and 5 step distance stimuli.

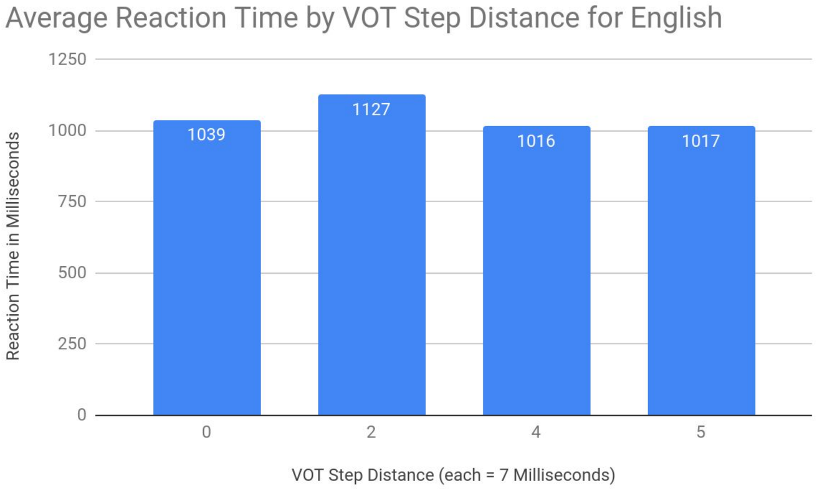

Table 6.

The pattern across the Russian conditions is much less clear. Stimuli utilized within the Russian discrimination experiment did not differ in VOT at a condistent interval. Thus analysis could only be conducted using absolute distance rather than step distance. Looking at the discriminability indices reported in Table 7, we observe much worse performance as compared to English discrimination. The only positive value being discriminability for acoustically identical stimuli. All other conditions are either negative or zero. The lowest discriminability is observed in the 10ms distance trials- these constituting the lowest non-acoustically identical distance stimuli pair. The discriminability of the remaining distance conditions do not appear to follow any tangible pattern. Of note however is that discriminability is best within these conditions at the farthest possible VOT distance between stimuli.

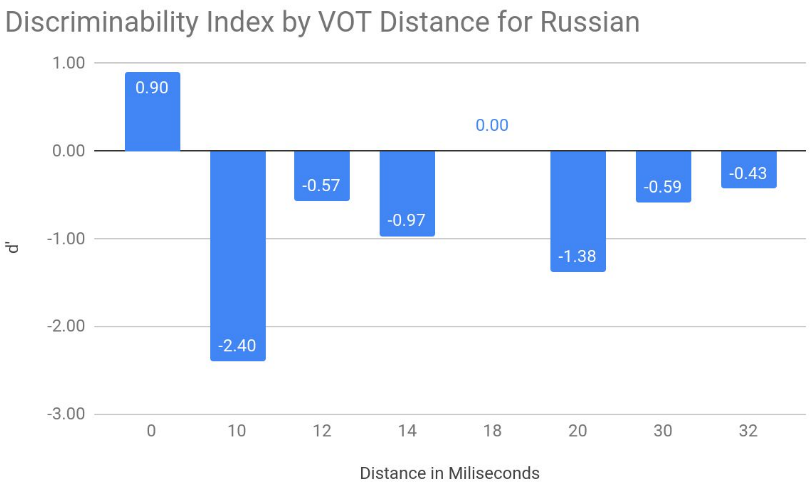

Table 7.

Reaction times seem to increase overall with VOT distance. Based off of Table 8 this pattern is not especially strong as there are exceptions where reaction times decrease at 18ms and 30ms distance. One point of note is that the fastest average reaction time observed is for the 0ms distance condition. The strong performance for identical stimuli and muddled poor performance for all other stimuli aligns fairly strongly with observed discriminability.

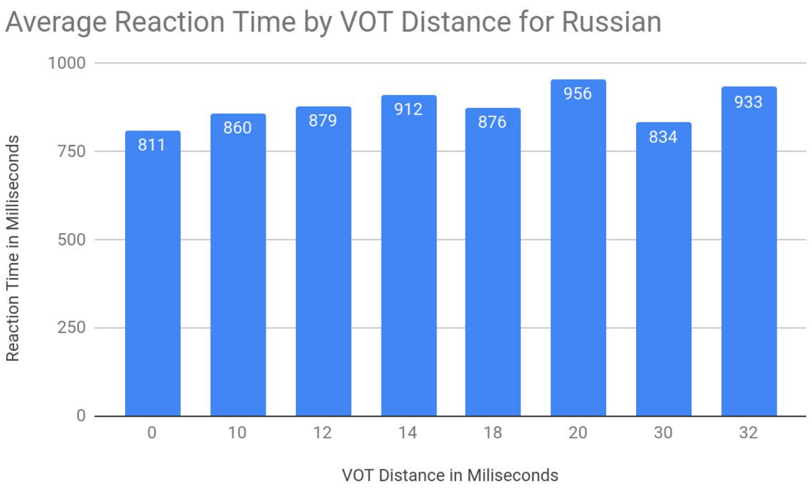

Table 8.

Discussion

There is a remarkable similarity between the results of Pisoni & Tash (1974) and the English data from both the identification and discrimination experiments. This is impressive given the lack of power for this small lab report. In Table 1 we observe nearly identical response percentages for stimuli ranging across the VOT continuum. And in Table 2 we observe a similarly unimodal distribution of reaction times around the locus of perceptual uncertainty. For discrimination, the higher reaciton times for 2 step conditions as compared with identical and longer distance conditions we see in Table 6 mirrors the pattern of reaction time Pisoni & Tash (1974) observe. The values are certainly not numerically identical, but the relationship between VOT step distances holds. The discriminability indices from Table 5 reveal this pattern as well. The signal for identical stimuli and distant stimuli is stronger than similar stimuli, this is predicted by the three stage model Pisoni & Tash (1974) put forth. All of this similarity lends some confidence to the interpretation of other results.

Within identification, the results from the Russian experiments seem to broadly fall within two distinct classes. One class of voicing lag and another of prevoicing. Looking at the percentage responses in Table 3 makes this clear. The two conditions with voicing lag, while not as consistent as the patterns from English identification, are predominately /ta/ identified. While any condition with prevoicing was predominately labelled as belonging to the /da/ category. This is interesting because prevoicing (at least of these consonants) is not part of the input English speakers receive. As well, the two conditions which were not prevoiced were at 2ms and 10ms of VOT and were still categorized as /ta/. The English identification results from Table 1 demonstrate that phones within this range are typically labelled as /da/ by English speakers. This shows that in spite of top-down grammatical experience, the participant was able to perform more similar to a typical Russian speaker by means of bottom-up acoustic information.

Switching to discrimination however, we observe the negative penalties of this conflicting top-down grammatical information. This is clearest in the contrast between the discriminability index values from English in Table 5 and Russian in Table 7. Performance for Russian is only positive when sounds are identical, while it does increase with distance the improvements are marginal. This weaker performance reflects the increase in noise in the signal due to conflicting grammatical knowledge and the unfamiliar phonetic quality of prevoicing.

A curious fact however is the difference in reaction times between English and Russian discrimination. While performance in Russian was worse, the reaction times to provide these responses was on average faster than for English. This can be seen in Table 8 and Table 6. The reason for this results is unclear. Due to the fact the Russian discrimination experiment run last, it it possible that some amount of fatigue could be at play. This contrast certainly demonstrates that reaction times are a meaningless measure within the analysis of this paradigm without some evaluation of accuracy. The discriminability index values demonstrate where performance is best across conditions.

One central motivation behind this research was to evaluate performance across languages when phonemes are shared, but the phones which describe those phonemes are different. This is in contrast with the research from Kazanina et al (2006) in which the Korean speakers lacked the same phonemes as Russian speakers, but the distribution of VOTs for the speech sounds was similar. In Kazanina et al (2006) Korean speakers appear to discriminate worst when the two speech sounds played are close in VOT and discriminate best when the two speech sounds are distant. The same appears true for the English speaking participant in Russian discrimnation. While the performance boost (as seen in Table 7) with VOT distance increases is not tremendous or consistent it is present. The lack of consistency and power of this effect could be on account of the unfamiliar prevoicining present in the Russian stimuli. Regardless it appears that this advantage which distance provides is somewhat independent of phonemic distinctions.

References

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Pisoni, D.B., Tash, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 15, 285–290 (1974). https://doi.org/10.3758/BF03213946

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(30), 11381–11386. https://doi.org/10.1073/pnas.0604821103