What, me saccade?

Abstract

Basically there is the vwp and it is used as a proxy for word recognition. This use follows from allopenna 1996, in which he showed the the proportion of fixations to different referrents matches what would be expected with TRACE (after suitable transformation, of course). This resulted in over two decades of vwp use for such purposes. In 2021, mcmurray asked if that curve really was what we thought it was. Through an analysis of generating hypotheses of increasing (but still relatively minimal) complexity via simulation, bob showed that even in cases of moderate complexity, we were not able to positively recover the generating curve responsible for eye movements. why was this, and what are the implications for understanding these curves? In this paper, we revist the 2021 princess bride paper and offer an explanation for the extreme demonstrated bias of the simulations. from this, we propose a new method for using vwp data to estimate underlying activation curves. finally, we top this all off by comparing the proposed method to allopena 1996 in a head-to-head, single elimination match up, a battle suitable for mount olympus itself in which two ultimate theories fight to a bloody and violent death to see which has been ordained by god to reign as total champion in the hearts and minds of language psychologists the world over.

1 Introduction

"introductions are the hardest part"

Spoken words create analog signals that are processed by the brain in real time. That is, as the spoken word unfolds, a cohort of possible resolutions are considered until the target word is recognized. The degree to which a particular candidate word is recognized is known as activation. An important part of this process involves not only correctly identifying the word but also eliminating competitors. For example, we might consider a discrete unfolding of the word "elephant" as "el-e-phant". At the onset of "el", a listener may activate a cohort of potential resolutions such as "elephant", "electricity", or "elder", all of which may be considered competitors. With the subsequent "el-e", words consistent with the received signal, such as "elephant" and "electricity" remain active competitors, while incompatible words, such as "elder", are

eliminated. Such is a rough description of this process, continuing until the ambiguity is resolved and a single word remains.

Our interest is in measuring the degree of activation of a target relative to competitors. Activation, however, is not measured directly, and we instead rely on what can be observed with eyetracking data, collected in the context of the Visual World Paradigm (VWP) (Tannenhaus 1995)[**?**].

Research goals: Some more context is helpful hear to try an understand what exactly it is that researchers are trying to learn from this data. As mentioned previously, we are largely concerned with "activation". In particular, though, we are often interested in how the activation of competing words compare. Specifically, we might ask, "at what point in the audio signal does the subject begin to identify the target word, relevant to competitors." This is of special interest in studying language development in typically developed children and those with cochlear implants, where atypical children may require more of a signal before they are able to disambiguate what they are hearing. As such, it is often of interest to ask when and how activation differs both between target words and competitors, as well as between different subjects. It is largely these last two areas that have dominated much of the VWP research.

Our specific goal here is to determine if a method of estimating activation, measured as being more consistent with the predictions of TRACE, can be improved by relying solely on saccade data rather than the traditional method using fixation based approaches

2 A brief history

start with a brief history to where we are today. I think the order goes:

1. TRACE in 1986 along with connectionist model of language

2. VWP by Tannenhaus 1995

3. VWP + TRACE, allopenna 1996

4. As far as I can tell, it's Bob's 2010 paper that was first to

   (a) Look at individual differences in word recognition (relevant for the "group distribution of curves" hypothesis and

   (b) Introduce parametric forms to be fit to the data (the assumption we continue to run with), or at very least, introduce ones that are interpretable

I can flesh these out in more details later, it's mostly narrative here and not crucial to the work being evaluated

TRACE   How speech is perceived and understood has been a subject of much debate for a significant portion of psycholinguistic's history. Starting in the 1980s and persisting today, most researchers subscribe to what is known as the connectionist model of speech perception. Briefly, this model posits that speech perception is best understood as a hierarchical dynamical system in which aspects of the model are either self reinforcing or self inhibiting with feedforward and feedback mechanisms. For example, hearing the phoneme \h\ as in "hit" will "feedfoward", cognitively activating words that begin with the \h\ sound. These activated words then "feedback" to the phoneme letter, inhibiting activation for competing phonemes such as \b\ or \t\. In 1986, McClelland and Elman introduced the TRACE[1] model implementing theoretical considerations into a computer model [**?**].

vwp   Tannenhaus 1995, really, though I don't want to go into more than just the basics, Magnuson 2019 has a good review of the history (amongst other things), but I'll leave this for now:

To briefly illustrate, the VWP is an experimental design in which participants undergo a series of trials to identify a spoken word. Typically, each trial has a single target word, along with multiple competitors. The target word is spoken, and participants are asked to identify and select an image on screen associated with the spoken word. Eye movements and fixations are recorded as this process unfolds, with the location of the participants' eyes serving as proxy for which words/images are being considered. Any particular trial could have anywhere from 3 to 10 eye movements, known as saccades, followed by fixations lasting between 200ms-100ms.

allopena   It was against simulated TRACE data that Allopenna (1998) found a tractable way of analyzing eye tracking data [**?**]. By coding the period of a fixation as a 0 or 1, depending on the referent and taking the average of fixations towards a referent at each time point, Allopenna was able to create a "fixation proportion" curve that largely reflected the shape and competitive dynamics of word activation suggested by TRACE. Since this publication, it has been this transformation of the data that has guided the last 25 years of VWP research. Could include here an illustration of VWP. also a lot of other various details from allopena

parametric methods and individual curves (sli paper)   Probably other developments in vwp world occured between 98 and here, but limited to what's relevant in this paper. In 2010, McMurray et al expanded the domain of vwp by introducing emphasis on individual differences in participant activation curves. While the implications of this are surely great in the application of the vwp, two aspects of this paper are relevant

---

[1]TRACE doesn't stand for anything – the name is a reference to "the trace", a network structure for dynamically processing things in memory

here. First, although they were not the first to introduce non-linear functions to be fit to observed data, they did introduce a number of important parametric functions in use today, namely the four (or five) parameter logistic and the double-gauss (asymetrical gauss). Second, which I suppose was also introduced by Mirman (2008) to some degree (though I have not read it yet) is specifying individual subject curves across participants. This has been critical along with the introduction of the four-parameter logistic in that:

1. The parameters of the functions describe interpretable properties
2. This made the idea of distributions of parameters for a particular group a relevant construct

As an aside, this also served as impetus for new methods for analyzing this data. A history of that is given in Seedorff 2018, as he describes other methods originally used prior to `bdots`, including AUC and cluster-based permutation testing. There is also some discussion of that in this paper, but again, not relevant to the task at hand.

At any rate, we now find ourselves in a place where TRACE-validated VWP data is being used to measure word recognition, and in a particular VWP study, subjects are measured, their data aggregated, and non-linear parametric curves with interpretable parameters are fit, giving rise to empirical distributions of curves across experimental or treatment groups. The comparison of these groups has given rise to bdots, which is significantly more badass now, see chapter 1 of my dissertation.

linking hypothesis? not sure if this is a section worth elaborating on. more interesting narrative, but relevant insomuch as our conclusions are adjacent to the conversations regarding linking hypotheses. Its been awhile since ive read magnuson's paper but i remember it being interesting. some minor inclusion of that here may make for an interesting call back/tie in when i bring up the saccade and its proximity to the construct (activation) in question.

3 Where we are now

Context in hand, we are ready to introduce some of the characters of our story. this includes the finer points of the vwp, eye tracking data, and how allopena's introduction ties in with bob's parametric

3.1 anatomy of eye movements

There are three components of eye movements with which we are concerned with here. The first two, saccades and fixations, are associated with physical mechanics of eye movements; the third, oculmotor delay, is a phenomenon related to the association between cognitive activaiton and physiological response. We will briefly introduce each of these topics here.
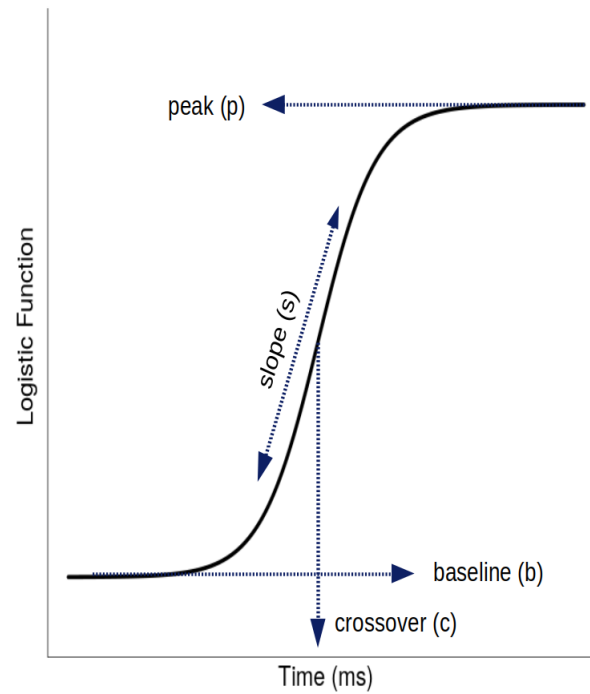
Figure 1: An illustartion of the four-parameter logistic and its associated parameters. also might be premature to include here since i haven't even shown the logistic or detailed its case, only its existence. Still, ill leave this up so that the image can be moved around if needed. What i would do in this caption otherwise is explain each of the parameters.
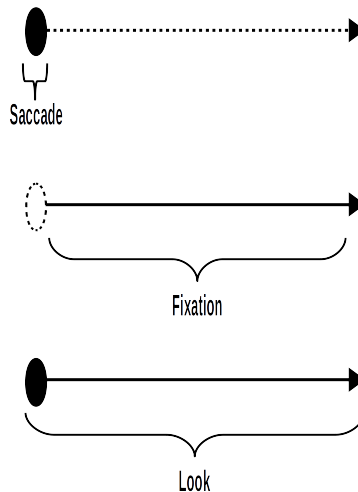
Figure 2: this obviously looks terrible here, need to redo image from odg file

Saccades and fixations: Rather than acting in a continuous sweeping motion, as our perceived vision might suggest, our eyes themselves move about in a series of short, ballistic movements, followed by brief periods of stagnation. These, respectively, are the saccades and fixations.

The short ballistic movements are known as saccades, periods of between 20ms-60ms (source? more accurate times?) in which they eye is in motion and during which time we are effectively blind. Once in motion, saccades have no ability to change their intended destination. Following the movement itself is a period of stillness known as a fixation, itself made up of a necessary refraction period from the saccade (time?) followed by a period of voluntary fixation; the typical duration of a fixation is (some length). Together, an initiating saccade and its subsequent fixation is known colloquially as a "look". See Figure 2.

Oculomotor delay: While the physiological responses are what we can measure, they themselves are not what we are interested in. Rather, we are interested in determining word activation, itself governing the cognitive mechanism fascillitating the movements in the eyes. It's suspected/stated/known (source?) that upon finishing a particular saccade, the mind is already anticipating where it will move next. What is relevant for our purpose here, however, is that the period of oculmotor delay is random, resulting in biased observations between what we are able to measure and what we are interested in discovering. How this phenomenon relates to saccades and fixations is demonstrated in Figure 3.

Alternatively, there is a full figure i could use here:

(thinking out loud here) The organization on this one is a little up in the air. It's interesting to include and will be one of the things we discuss in more detail. it seems a bit premature to bring in the whole idea of ρ *here* when that might be better suited to preparation for the simulation/comparison section.
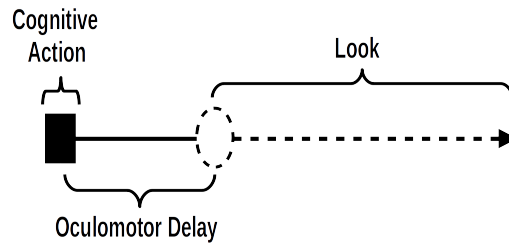
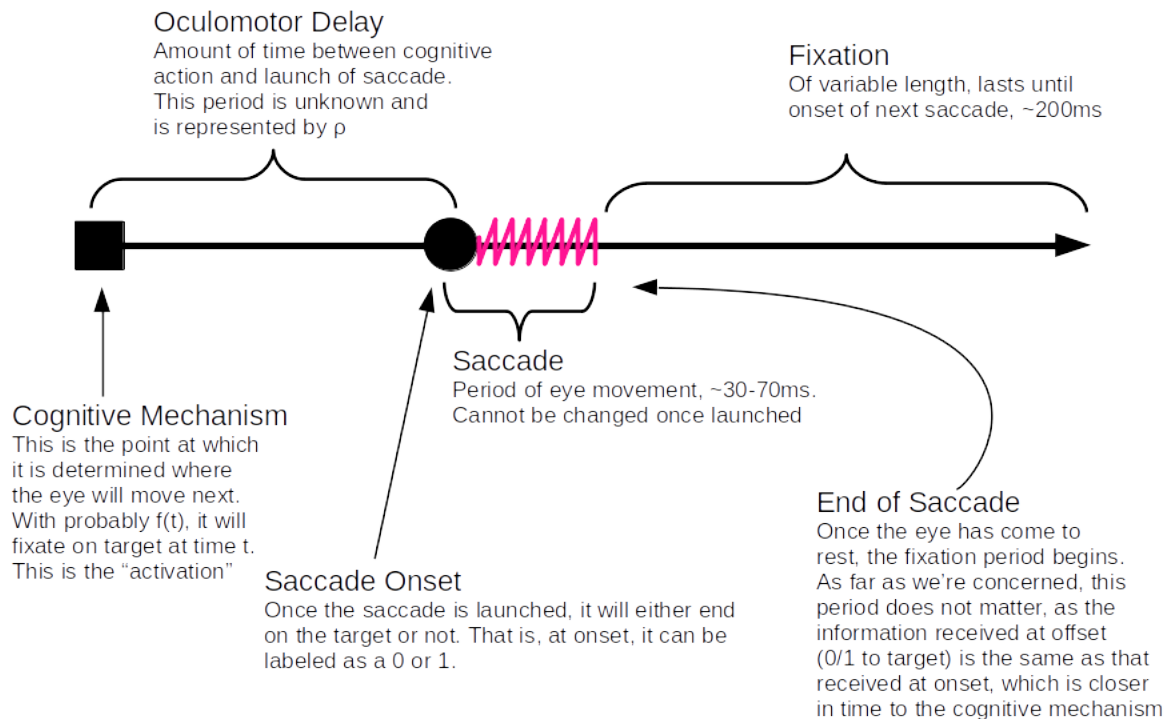Figure 3: this also could probably be reformatted or made bigger

Oculomotor Delay
Amount of time between cognitive action and launch of saccade. This period is unknown and is represented by ρ

Fixation
Of variable length, lasts until onset of next saccade, ~200ms

Cognitive Mechanism
This is the point at which it is determined where the eye will move next. With probably f(t), it will fixate on target at time t. This is the "activation"

Saccade
Period of eye movement, ~30-70ms. Cannot be changed once launched

Saccade Onset
Once the saccade is launched, it will either end on the target or not. That is, at onset, it can be labeled as a 0 or 1.

End of Saccade
Once the eye has come to rest, the fixation period begins. As far as we're concerned, this period does not matter, as the information received at offset (0/1 to target) is the same as that received at onset, which is closer in time to the cognitive mechanism

Figure 4: This figure actually doesn't look too bad, but may be better when articulating how saccades measured and why (also includes info on $f(t)$, ρ, etc., so maybe we will present this later around the time of simulation

7

3.2 VWP data

technically this kind of includes the "proportion methd" too, not sure if i should break up or how exactly to treat this

We now consider how the aforementioned mechanics relate to the VWP. In a typical instantiation of the VWP, a participant is asked to complete a series of trials, during each of which they are presented with a number of competing images on screen (typically four). A verbal cue is given, and the participants are asked to select the image corresponding to the spoken word.

An individual trial of the VWP may be short, lasting anywhere from 1000ms to 2500ms before the correct image is selected. Prior to this, the participants eyes scan the environment, considering images as potential candidates to the spoken word. As this process unfolds, a snapshot of the eye is taken at a series of discrete steps (typically every 4ms) indicating where on the screen the participant is fixated. While there is evidence of cognition happening behind the scenes in a continuous fashion (spivey, mouse trials), an individual trial of the vwp may contain no more than four to eight total "looks" before the correct image is clicked, resulting in a paucity of data in any given trial.

To create a visual summary of this process aggregated over all of the trials, a la Allopena, a "proportion of fixations" curve is created, aggregating at each discrete timepoint the average of indicators indicating that a participant is fixated on a particular image. A resulting curve is created for each of the competing categories (target, cohort, rhyme, unrelated), creating an empirical estimate of the activation curve, $f_\theta(t)$. See Figure 5. Mathematically, it looks like this:

$$y_{it} = \frac{1}{J} \sum z_{ijt} \tag{1}$$

where z_{ijt} is an indicator $\{0,1\}$ for subject i in trial j at time t and such that

$$f_\theta(t) \equiv y_t. \tag{2}$$

Nonlinear parametric curves can be fit to the observed y_t as is done in the `bdots` package in R (see chapter 1).

Here, I've copied and pasted what I had about this elsewhere. There are bits and pieces I like from each, so for now going to leave them both and make them "do it" and have babies later.

One method employed to use this data involves measuring intervals of fixation to a target over a series of trials. For each trial i and time point t (typically sampled at intervals of 4ms, i.e., $t = 0, 4, \ldots, 2000$), we collect a sample of z_{it}, an indicator of whether a participant is fixated on the target object at that point in
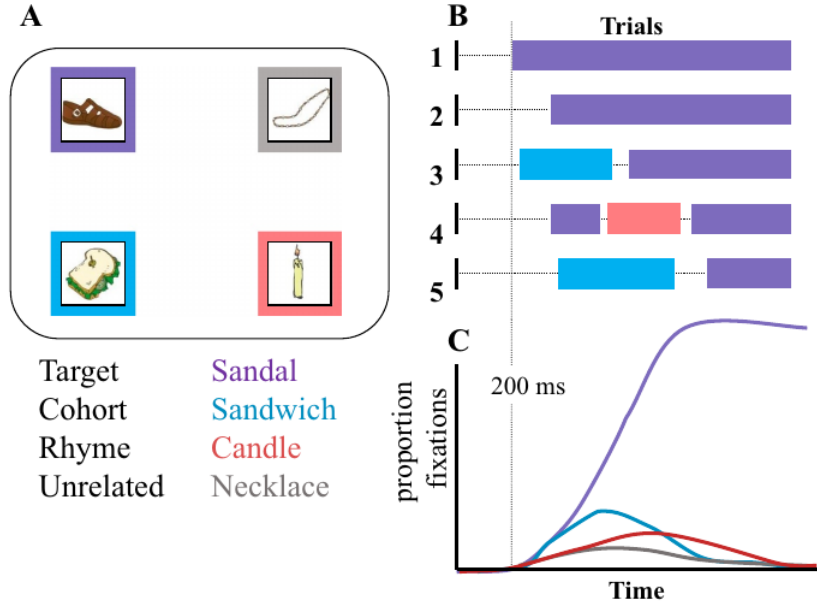
Figure 5: This screenshotted from Bob's princess bridge paper. i would like to reconstruct a similar illustration here as it does a great job illustrating the point. *However*, this section as it stands may make more sense elaborated elsewhere, in particular where I give a mathematical treatment to what the "fixation curve" is

time. Averaging over the collection of trials, we construct an estimate of $f_\theta(t)$,

$$y_t = \frac{1}{J} \sum_i z_{it}.$$

In other words, it is implicitly assumed that the trajectory of the eye follows the trajectory of activation, where the average proportion of fixations at a particular time is a direct estimate of activation. As each individual trial is only made up of a few ballistic movements, the aggregation across trials allows for these otherwise discrete measurements to more closely represent a continuous curve. Curve fitting methods, such as those employed by 'bdots', are then used to construct estimates of function parameters fitted to this curve.

[this really belongs with a compare-and-contrast section following saccades. primary benefit of this....over what?] One of the primary benefits of this method is that it captures the duration of fixations, with longer times being associated with stronger activations. This becomes important when differentiating fixations associated with searching patterns (i.e., what images exist on screen?) against those associated with consideration (is this the image I've just heard?). A shortcoming, however, is that it conflates two distinct types of data, generated via different mechanisms, the fixation and saccade.

9

4 where we are going

4.1 princess bride paper

From the abstract of this paper: "All theoretical and statistical approaches make the tacit assumption that the time course of fixations is closely related to the udnerlying activation in the system. However, given the serial nature of fixations and their long refractoryh period, it is unclear how closely the observed dynamics of the fixation curves are actually coupled to the underlying dynamics of activation."

This is a critical statement to be made, and in a sense, it ties into the general idea of the linking hypothesis presented in Magnuson. I make no attempts here to challenge or otherwise present a drastic new linking hypothesis. Rather, I take assumption made here: an outline (because i love enumerated lists) of the events, in order, is as follows:

1. Some measure of activation is occuring in the mind as a consequence of audio stimuli
2. Through some indeterminate mechanism (linking hypothesis), this level of activation governs the mechanics of eye movements, specifically the direction/location of saccades and duration of fixation
3. The resulting mechanism governing *saccades*, namely, where to look *at a specific period in time* follows a sigmoid curve, here specifically the four-parmeter logistic

n.b., bob refers to these as "fixation curve"

In 2022, McMurray brough into question the validitity of a standard VWP analysis, and a more thorough treatment of his presented arguments is warranted but for the time being counts as narrative and so the elaboration will wait. For now, we will present those elements that are crucial for understanding the direction of the methodology to be presented.

In short, the question that is being gotten at is: are we able to recover the underlying dynamics of the system in question (activation) in light of the "nature of the fixation record as a stochastic series of discrete and fairly long lasting physiologicall constrained events?" In short, the answer is no. McMurray notes that the typical, unspoken assumption implicit in VWP studies is the "high frequency sampling" (HFS) assumption, which states that the underlying activation at some time determines the probability of fixation. He then goes on to note that this is "patently" untrue and is nothing more than a polite fiction.

Nonetheless, it is useful to compare the relationship of the underlying dynamics (as we will elaborate upon further, a generating function) with the observed data in the context of the HFS, relative to other, more complex assumptions. This is done through a series of simulations, each with their own set of stochastic mechanics determining eye movements and fixations. In all cases, however, it is assumed that there is exists an underlying generating function that, at any particular time, is responsible for dictating some aspect of

a subsequent fixation. We will start with an overview of the general algorithm for an individual subject, followed by a brief summary of each of the simulations.

Algorithm:

1. A set of generating parameters for the four-parameter logistic is drawn from an empirically determined distribution. This curve, $f_\theta(t)$, is treated as the probability of fixating on a target at time t

2. After a random offset start time, t_0, a binomial random event is drawn determining the probability of fixating on the target, $p \sim Bin(f_\theta(t_1))$

3. After this initial draw, a fixation occurs for a period of time:

   (a) Under the HFS assumption, this period is instantaneous – that is, whatever the time, t is also the probability of fixation

   (b) Under the FBS assumption, the length of the fixation draws from a gamma distribution, ending at time time t_2

   (c) Under the FBS+T assumption, again a random length fixation is drawn from a gamma distribution, but with a higher mean value if the fixation is drawn to the target

4. idk im desribing this weird – maybe come back to this list later

what I can do instead is offer a brief written summary of each of the methods.

High Frequency Sampling (HFS)   The underlying activation of the word *is* the probability of fixating at a particular time.

Frequency Based Sampling (FBS)   The FBS assumption differs from that of the HFS assumption in that the observed data is gathered from a period of fixations of random duration. Once each fixation is "drawn", the subject remains fixated on a particular object for the full length of the fixation. The next fixation's location is determined at the *onset* of the previous fixation. In particular, this simulation assumes that immediately once a fixation is made, the subject begin's preparing to launch their next saccade

Frequency Based Sampling + Target (FBS+T)   This simulation is identical to the previous with the exception that duration of fixations to the target, while still random, follow a different different distribution than fixations to non-targets, with longer durations afforded to target fixations to account for "information gathering behavior".

Naturally, as the complexity of the assumptions increased, so did the observed bias in recovering the parameters of the generating function.

At any rate, I'm getting to caught up in the particulars when what I really want or need to say is quite simple. It comes down to this: *the only observed behavior governed by the generating curve is the saccade when launched.*

saccades   The entirety of the bias resulting from FBS and FBS+T we the consequences of two facts, or put differently, there are two sources of bias that we need to consider:

1. We were "observing" data points $\{0, 1\}$ at any time t without having observed any behavior from the generating curve at that time (not sure what to call this, added observation bias?)

2. When we did sample directly from the curve at fixation onset, we were actually sampling from the onset of the previous fixation (delay bias)

The solution to this, then, is to simply use the saccade data, or only collect as $\{0, 1\}$ the instance at which a fixation occurs, discarding the rest. While this does not address the delay bias, it does remove a significant amount of bias from the added observations. One obvious shortcoming is that is dismisses all "information gathering behavior" that could otherwise be gleaned from the duration of fixations. To what effect this or other enhancements may have on the efficiency of this data are yet to be seen, but at very least it offers a more clearly defensible relationship between the observed data and the generating function.

The idea of information gathering behavior is a useful one, but it assumes a linear relationship between the length of time of the fixation and strength of activation. However, one might suspect that after a period of necessary refraction, each subsequent period of time gives exponentially more weight to the argument of activation. A potential consequence of this is that an indication of fixation 50ms following the launch of a saccade may convery different information than the indication of a fixation still present 300ms following a saccade, despite the fact that these are recorded equally as $\{0, 1\}$. That is, under the present system, rather than indicating the gathering of more information, longer fixations simply increase both the bias due to added observation *and* the amount of bias on account of delay (as the subsequent fixation will have been determined further removed from its occurence when following a longer fixation). On the other hand, a mechanism for recording information-gathering behavior may be more readily implemented in a saccade-style method whereby each saccade is weighted by the length of its subsequent activation, for example.

———

From here, we will describe the proposed saccade method in more detail, compare the results of using saccade data against what was found in mcmurray, see that it largely resolves the problems, and then ask the natural question: how does this stack up against what allopenna found?

Also note somewhere that we are primarily limiting discussion to the logistic here. A brief treatise on the asymetric gaussian is given in the appendix
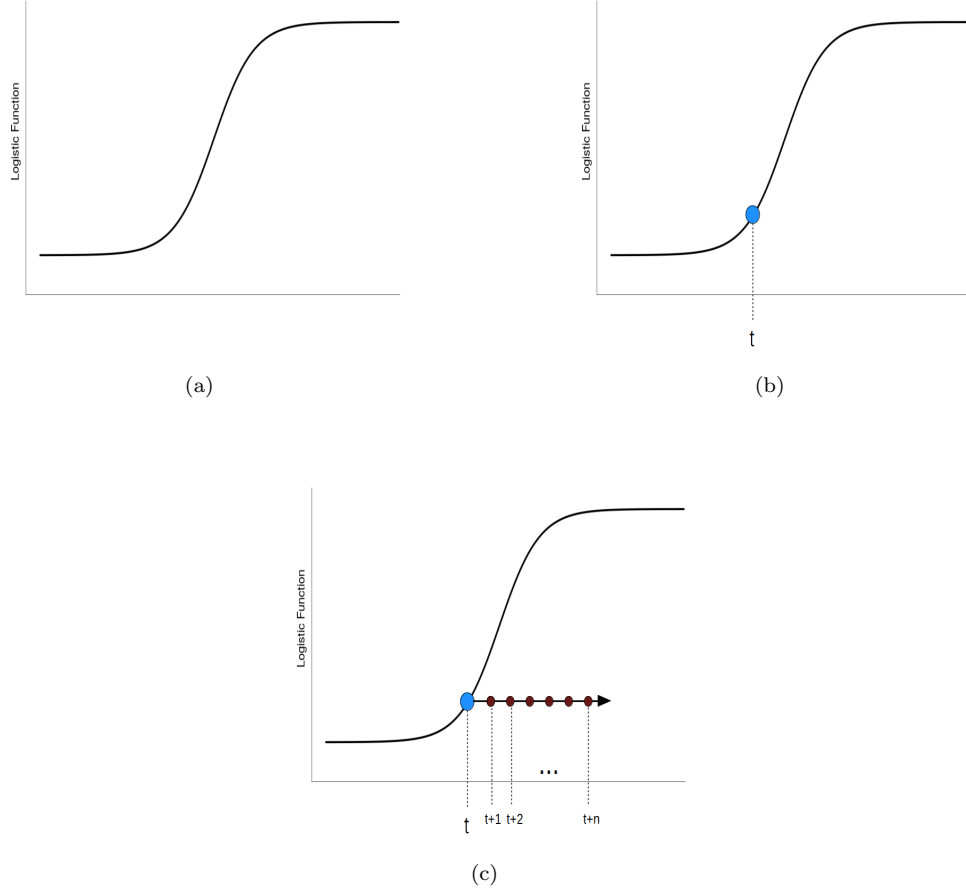
Figure 6: These illustrations can all be made larger (they were made for slides in an image editing program), but they illustrate the main point. **(a.)** here we see an example of a generating logistic function **(b.)** at some time, t, a saccade is launched (in the algorithm, a binomial is drawn with probability $Bin(f_\theta(t))$ **(c.)** at subsequent times, $t+1, \ldots, t+n$, we are recording "observed" data, adding to the proportion of fixations at each time but without having gathered any additional observed data at $f_\theta(t+1), \ldots, f_\theta(t+n)$, thus inflating (or in the case of a monotonically increasing function like the logsitic, deflating) the true probability.

4.2 Saccade method

If we are to consider eyetracking data samples from some probabalistic curve, it becomes necessary to differentiate between the two types. A saccade launched at some time, t, can be considered a sample from a data-generating mechanism at t. The duration of time between a given saccade and the one following follows a different mechanism altogether. By clearly delineating the mechanism from which we are sampling, we are able to reduce observed bias in the reconstruction of the activation curve.

In light of this, and in contrast to the fixation method, we propose estimating the activation curve with the saccade data alone. The primary benefit of this is two-fold. First, as suggested above, by decoupling two different types of data we are able to be more precise in what it is we are sampling. Second, as explained in the previous section, we are removing the added observation bias.

An important difference between these two methods is in the structure of the data itself. Whereas the former collects an array of data, with an observation for each time point in each trial, the saccade method is sparse, with the observed data indicating the outcome of the saccade, as well as the time observed. It is best represented as a set of ordered pairs, $\mathcal{S} = \{(s_j, t_j)\}$, with j indexing each of the observed saccades, and with

$$s_j \sim Bern(f_\theta(t_j)). \tag{3}$$

A value of $s_j = 1$ indicates a saccade resulting in a fixation on the target.

As with the proportion method, the observed data can be used as input for `bdots` to construct estimates of generating parameters.

5   Simulations Against Princess Bride

Here we now replicate the results of the princess bride paper, though with a few adjustments in light of the previous discussion. In particular, we noted that the two types of bias presented in the original simulations were added observation bias and delay bias. The first, added observation, we are addressing with the propsed saccade method. As to the second, we first elaborate here with a brief discussion of the delay bias and how it may also related to oculomotor delay.

In the original princess bride paper, FBS and FBS+T were only differentiated by the amount of additional bias introduced in FBS+T. Specifically, by drawing fixations to the target from a gamma distribution with a larger mean, we were both increasing the amount of added observation and delay bias, both consequence of the longer fixation period. Understanding that these differ in degree rather than kind, we collapse them

into a single construct here, as we will elaborate on shortly.

We also make adjustments to how we deal with oculomotor delay. As was shown in the simulations with the HFS assumption, a fixed oculomotor delay simply resulted in a horizontal shift of the estimated function, having otherwise no impact on the *shape* of the function. In contrast, added observation and delay bias both drastically impact the final shape. Further, in typical instances in which we are using the VWP, we are more frequently concerned with the relative difference between two curves rather than the curves themselves. The magnitude and relative location of such differences will be preserved under a horizontal shift, having ultimately little consequence on the resulting analysis.

In light of this, we will seek to combine the functional impact of oculomotor delay with the impact of more complex eye behavior in the following way: we recognize that with the exception of the HFS assumption (which is not considered here), any observation at time t will have been prompted at some time previously (that is, drawn from the activation curve). The length of this delay will be denoted $\rho(t)$. In the case of the HFS assumption, for example, this simply would have been $\rho(t) = 200ms$. For the FBS and FBS+T case, it would have been $\rho(t) = 200ms + $ length of previous fixation. As such, we can reduce the conditions under which we compare the fixation and saccade methods to two scenarios:

1. $\rho(t)$ is a constant function, including zero
2. $\rho(t)$ is a random variable, independent of the value of t_j

As in the princess bride paper, we will let $f_\theta(t)$ be a four parameter logistic, representing of generating or activation curve. Understanding that what we observe at time t was drawn from this function at time $t - \rho(t)$, we will differentiate the underlying activation curve the the observed data,

$$g_\theta(t) = f_\theta(t - \rho(t)) \tag{4}$$

Each simulation will be conducted with $N = 300$ trials, sampled from the same data generating function for each, with the attempted recovery of the generating curve done using the `bdots` package.

5.1 Known Delay

In the case in which the oculomotor delay is known, an unbiased recovery of the data generating curve is not an issue – we simply horizontally shift each observed saccade by its known oculmotor delay.

5.2 Unknown Fixed Delay

The simulation was conducted using a fixed oculomotor delay of $\rho = 200ms$. Although the resulting recovered curve is biased, this bias simply results in a horizontal shift, $g(t) = f(t - \rho)$. This is especially relevant in a situation in which we are interested in comparing the data generating curve between two groups.

For example, one method of analyzing VWP data (which inspired the 'bdots' package) was to determine on which intervals $I = \cup_k I_k$ two data generating curves were statistically different. Suppose, for simplicity, that there is an interval $I = [t_1, t_2]$ on which the difference between two curves, $f(t|\theta_1) - f(t|\theta_2)$, is statistically significant. Given that we observe $g_i(t) = f(t - \rho|\theta_i)$, we would simply find that a significant difference occurs at $I + \{\rho\} = [t_1 + \rho, t_2 + \rho]$, a horizontal shift resulting from the oculomotor delay.

In other words, the size of the interval would remain the same, and the relative differences between curves would be preserved under a horizontal shift.

5.3 Unknown Random Delay

The final scenario for consideration involves a situation is which the OM delay is unknown and random. Here, the bias will not be resolved with a simple horizontal shift, and the shape of the curve itself may be different between the one generating the data and the one observed. This has largest implications when comparing estimated curves between two groups.

We will interogate a number of potential methods for dealing with this issue, though we do feel confident that, even with this known bias, our proposed method will still be preferable to existing ones.

6 Compare with TRACE

This section needs to move to the prominent spot since its the meat of my argument Here is a snippet I stole from elsewhere (that i also wrote):

———————

In Allopenna (1998), they say "... A time course analysis of the proportion of fixations on the target object (currect referent), cohort competitor, and unrelated objects suggested that this measure would be extremely sensistive to the uptake of information during the lexical process. *Futhermore, the shapes of the functions suggested that they could be closely mapped onto activation levels*...*We also put forth an explicit account of the mapping between activation levels simulated by the TRACE model and fixation probabilities.*"

That is, it seems like the connection between fixations and activation was first posited due to the simi-

larities in shape of the functions with TRACE models. ("linking hypothesis" – also, see magnuson)

Actually, this entire paper gives an explicit transformation from what was predicted in TRACE to what should be predicted as proportion/probability of fixation

---

6.1 Collecting TRACE data

Here, I want to cover a handful of things:

1. Where did this trace data come from (bob's paper)

2. How could it be reconstructed in trace (parameters)

3. What words were used

4. How did I get it into acceptable format for me

5. Luce choice rule, time transformation, scaling term

We validate our arguments above using a fortuituous study by McMurray (2010) in which subject data was collected to be analyzed against a collection of hyperparameters in TRACE testing a number of theoretical constructs in language impairment, including sensory and phonological confusion, vocabulary size, etc., presenting us with both a collection of empirically collected data from the subjects, as well as accompanying TRACE data, simulated with words used in the empirical portion of the study. Using this, we will be able to examine the relationship between empirically collected look data, saccade data, and validating TRACE data.

6.2 TRACE Data

In McMurray (2011), a range of simulated TRACE data was collected across the space of hyperparameters, each simulation containing 14 trials, each with a Target, Cohort, Rhyme, and Unrelatd object (Appendix B. in Bob's paper). Here, we included only TRACE data associated with the default parameters specified in jTRACE (magnuson) and found the average TRACE activation across frames/cycles for each object (target, cohort, etc.,).

As has been discussed elsewhere (Allopena, Magnuson, Bob), TRACE outputs a set of activations relative to all of the words included in its lexicon, and what is needed is a linking function the is able to return a simple mapping of TRACE activations to an estimated probability. We follow the steps given in the appendix of McMurray (2011), which we briefly outline here.

First is an implementation of the Luce Choice Rule (from somewhere) which dictates that, when considering the probability of fixation of a candidate word, we need only consider the probability of fixation relative to the other candidates present rather than our entire lexicon. This means that in the context of the VWP, our probability of fixation would be relative to the (usually three) other objects on the screen, rather than the thousands of words with which we may be familiar.

Really, though, does this have to be done at all? Can I not simply point to bob's paper and be like, look, this is the trace data and this has all the data on the subjects. here we are just going to look at the results because nothing else matters or is relevant. Or this can be in the appendix

7 Discussion

what have we learned?

8 limitations

probably good idea to keep running list of these all in one place

1. linking hypothesis/cognition curve

2. trace parameters maybe/general degrees of freedom

3. only evidenced on logistic, though for practical not theoretical reasons

4. adding parametric form (necessity for saccade method)

5. oculomotor delay, where to discuss

9 appendices

Here I am just including more or less random sections that either do not have a definite place yet in the main body of the paper, are part of what might be considered future work, or truly are things that belong in the appendix

Appendix A

Here, I can talk about preparing the TRACE data from bob's paper, as well as preparing the other data. In particular, I can include info on changing the temperature parameter from luce choice rule

Appendix B

Here, I can discuss parallel results when I adjusts for reaction time when gathering observations

Appendix C

Oculomotor Delay We begin with an assumption that the curve of interest can be represented parametrically. For example, the four parameter logistic, defined as

$$f(t|\theta) = \frac{h - b}{1 + \exp\left(4 \cdot \frac{s}{h-b}(x - t)\right)} + b,$$

is often used to describe the trajectory of probability of a subject launching a saccade and fixating on the target location while simultaneously used as a proxy for word activation. To illustrate, a subject with the depicted fixation curve may initiate a saccade beginning at time $t = 970$, with a probability of $p = 0.5$ of subsequently resting on the target:

[insert image]

Mentioned previously, this same parametric curve is used in both the proportion and saccade methods.

As saccades are easily gathered from available eyetracking data, we are, in principle, able to collect samples directly from this curve. This goal is complicated, however, by oculomotor delay. That is, an observed saccade at t_j is likely a sample from the fixation curve $f_\theta(t)$ at some point prior to t_j. The degree to which this delay occurs, as well as the between and within subject variability of this delay, is a matter of active investigation. Most generally, we may consider an observation s_j at time t_j to be distributed

$$s_j \sim Bern[f_\theta(t_j - \rho(t_j))],$$

where $\rho(t)$ represents oculomotor delay. As written, we may consider circumstances in which:

1. $\rho(t)$ is a constant function (including 0)

2. $\rho(t)$ is a random variable, independent of the value of t_j

3. $\rho(t)$ is a random variable, dependent on t_j and possibly other aspects of the trial

To differentiate between the underlying data-generating mechanism and what is observed, we let

$$g_\theta(t) = f_\theta(t - \rho(t)),$$

where $g_\theta(t)$ is what is *observed* at time t. A saccade planned at $t = 300ms$ with an oculomotor delay of $\rho = 200ms$ will be observed at $t = 500ms$. That is, $g_\theta(500) = f_\theta(500 - 200) = f_\theta(300)$.

At present, it is common under the proportion method to account for this delay via a 200ms shift of the entire constructed proportion curve. We will propose instead a method whereby each saccade may be shifted individually and less homogenously. Reasons and implications for this will be presented in the next section.

We now consider a variety of scenarios for oculomotor delay and the subsequent impacts on the recovery of the underlying fixation curve from the observed data.

Appendix D

Idk, maybe oculmotor delay? That is kind of a disjoint section of this paper, and i don't think would get more than a mention, really, outside of a published paper. but its relevant for future direcitons, it's relevant for calling things what they are, it's not relevant for considering trace vs empirical results. Here is copy pasted out of the discussion of old paper what I had originally recordered for this:

With regards to the curve described, there are a number of avenues seemingly worthy of investigation. The most pressing of these appears to be methods to minimize the amount of bias present in scenario three, which presents the largest obstacle in the functional recovery of the data generating mechanism. Of special note here is the fact that the particular intervals in which this bias occurs can have a large effect on the overall bias, over and above that introduced by the occulomotor delay.

For example, consider the plot above in the situation in which there is an unknown random delay (blue curve). We may observe at 500ms the value of the data generating mechanism at 300ms (that is, we observed $g_\theta(500) = f_\theta(300)$), while $g_\theta(500) \approx f_\theta(500)$. In other words, the bias over this area is small if we make no correction.

In contrast, an observation at $t = 1000$ results in a highly biased estimate, as $g_\theta(1000) \ll f_\theta(1000)$. Accordingly, we note that the amount of bias at an observed point is a function of the derivative of the data generating function in a neighborhood of that point. Whether or not this observation proves profitable remains to be seen.

There also seems to be value in finding a way to incorporate the length of fixations into the modeling process. For example, we might consider the impact of weighting each saccade by the duration of its subsequent fixation, as it seems intuitive that saccades resulting in longer fixation periods are more likely initiated by activation rather than a searching pattern.

Appendix E

Why no double gauss treatment? simply for the reason that it is unable to find optimal fits in bdots curvefitters. in other words, with double gauss as it relates to the saccade method, there is a shortcoming in implementation rather than anything unique to it theoretically. as this is a theoretical paper, we limit presentation to those methods that may be implemented successfully on a computer as well.