

What, me saccade?

Abstract

Basically there is the VWP and it is used as a proxy measure for word recognition. This use follows from Allopenna 1996, in which he showed the the proportion of fixations to different referents 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 revisit the 2021 princess bride paper and offer an explanation for the demonstrated bias of the simulations. from this, we propose a new method for using VWP data to estimate underlying activation curves. We conclude by demonstrating consistency of the proposed method with existing continuous mapping models of word activation.

Notes:

1. Sections that are more “narrative” are less fleshed out. This includes VWP, TRACE, general history, etc., Some of these sections I just said what I would say
2. Citations are hard coded in here awaiting a bib to be created
3. Some plots/graphics need to be redone for size
4. There is some meta commentary, partially for the reader, mostly for me

1 Introduction

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 eye-tracking data, collected in the context of the Visual World Paradigm (VWP) (Tannenhaus 1995). In the last few years, researchers have begun to reexamine some of the underlying assumptions associated with the VWP, calling into question the validity or interpretation of current methods. We present here a brief history of word recognition in the context of the VWP, along with an examination of contemporary concerns. We address some of these concerns directly, presenting an alternate method for relating eye-tracking data to lexical activation. Finally, we show consistency of our method with existing continuous mapping models of activation using empirical data along with predictions made by TRACE.

This section needs work but it mostly covers the gist of what I am trying to convey, namely we are about to go from history → current state of the world → proposal and comparison → validation.

2 A brief history

We begin with a brief history to give context to later discussion. In particular, we will consider one of the leading theoretical models in speech perception, TRACE, followed by the introduction of the leading experimental paradigm, the VWP. We examine empirical evidence for the relation between these, and relevant theoretical advancements that have been made. Topics here are presented only briefly and limited to those directly relevant to the present work. For a fuller discussion of the history and uses of VWP, use google. (Or Huettig 2011b?)

An outline of the presentation (for internal use only):

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 among first to
 - (a) Look at individual differences in word recognition (not counting the ortho polynomial fits) (also 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

All of the paragraphs in this section are narrative and not mission critical. Need to be fleshed out

TRACE How speech is perceived and understood has been a subject of much debate for a significant portion of psycholinguistics’ history. Starting in the 1980s and persisting today, many 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 feed-forward and feedback mechanisms. For example, hearing the phoneme $\backslash h \backslash$ as in “hit” will “feed-forward”, cognitively activating words that begin with the $\backslash h \backslash$ sound. These activated words then “feedback” to the phoneme letter, inhibiting activation for competing phonemes such as $\backslash b \backslash$ or $\backslash t \backslash$. In 1986, McClelland and Elman introduced the TRACE¹ model implementing theoretical considerations into a computer model [?]. Maybe useful here to discuss activation, sigmoidal shape, etc.,

VWP 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.

Relating TRACE to VWP 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 for each 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, both for the target object, as well as competitors. This also served to establish a simple linking hypothesis, specifically, “We made the general assumption that the probability of initiating an eye movement to fixate on a target object o at time t is a direct function of the probability that o is the target given the speech input and where the probability of fixating o is determined by the activation level of its lexical entry relative to the activation of other potential targets.” Further of note is what this linking hypothesis does not include, namely:

1. No assumption that scanning patterns in and of themselves reveal underlying cognitive processes
2. No assumption that the fixation location at time t necessarily reveals where attention is directed (only probabilistically related to attention)

Other assumptions included here include that language processing proceeds independent of vision (Magnuson 2019), and that visual objects are not automatically activated. Or, more succinctly, it assumes that

¹TRACE doesn’t stand for anything – the name is a reference to “the trace”, a network structure for dynamically processing things in memory

fixation proportions over time provide an essentially direct index of lexical activation, whereby the probability of fixating an object increases as the likelihood that it has been referred to increases.

While other linking hypotheses have been presented (Magnuson 2019), that there is *some* link between the function of fixation proportions and activation has guided the last 25 years of VWP research.

Parametric Methods and Individual Curves While there have most certainly been advancements to the use of the VWP for speech perception and recognition (and expanded into related domains, such as sentence processing and characterizing language disorders (according to Bob)), we limit ourselves here to one in particular. In 2010, McMurray et al expanded the domain of the VWP by introducing emphasis on individual differences in participant activation curves. Two aspects of this paper are relevant 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 (asymmetrical gauss), the primary benefit being that the parameters of these functions are interpretable, that is, they “describe readily observable properties.” Second, which I suppose was also introduced by Mirman (2008) to some degree (though I have not read it yet, just pulling from Bob) is specifying individual subject curves across participants. This has been critical 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

Though not stated directly (given it predates `bdots` by 8 years), this also served as the impetus for investigating group differences in word activation through the use of bootstrapped differences in time series (Oleson 2017) and the subsequent development of the `bdots` software in R for analyzing such differences. (A history of exploring differences in group curves can be found in (Seedorff 2018)).

This brings us to the current day, where the state of things is such that TRACE-validated VWP data is widely used to measure word recognition by collecting data on individual subjects and fitting to them non-linear parametric curves with interpretable parameters. Context in hand, we are now able to introduce some of the main characters of our story, specifically how data in the VWP is understood and used.

3 Where we are now

This section includes the finer points of the VWP, eye tracking data, and how `allopna`’s introduction ties in with `bob`’s parametric proposition.

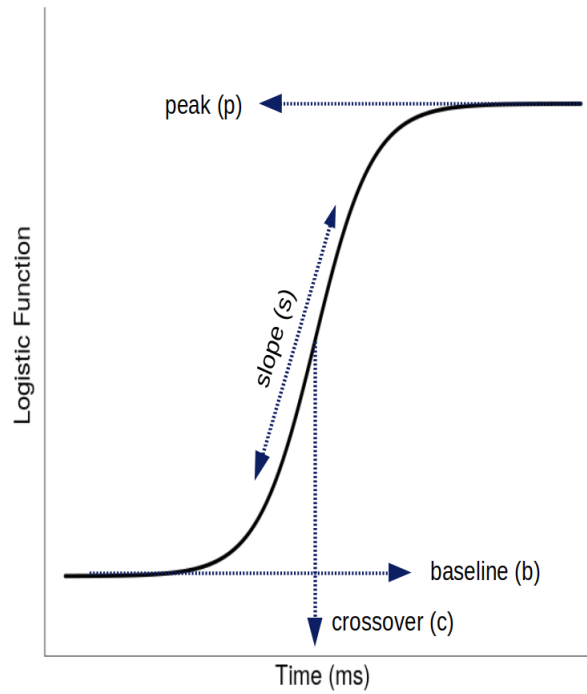


Figure 1: An illustration of the four-parameter logistic and its associated parameters, introduced as a parametric function for fixations to target objects in McMurray 2010. Can describe the parameters in detail, but should also have the formula itself somewhere to be referenced. (Equation 1)

3.1 anatomy of eye movements

There are three components of eye movements with which we are concerned. The first two, saccades and fixations, are associated with physical mechanics of eye movements; the third, oculomotor delay, is a phenomenon related to the association between cognitive activation and physiological response. We will briefly introduce each of these topics.

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 the 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?)

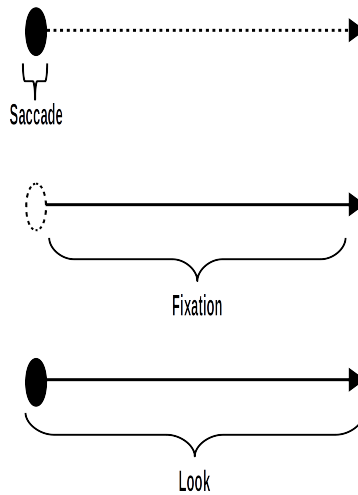


Figure 2: This image needs to be recreated for size. Illustrates saccade, fixation, and look

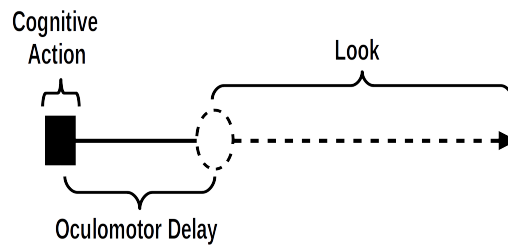


Figure 3: this also could probably be reformatted or made bigger

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 are not themselves what we are interested in. Rather, we are interested in determining word activation, itself governing the cognitive mechanism facilitating 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. Length of about 200ms also thrown around a lot. What is relevant for our purpose here, however, is that the period of oculomotor delay is a (likely) random process, 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.

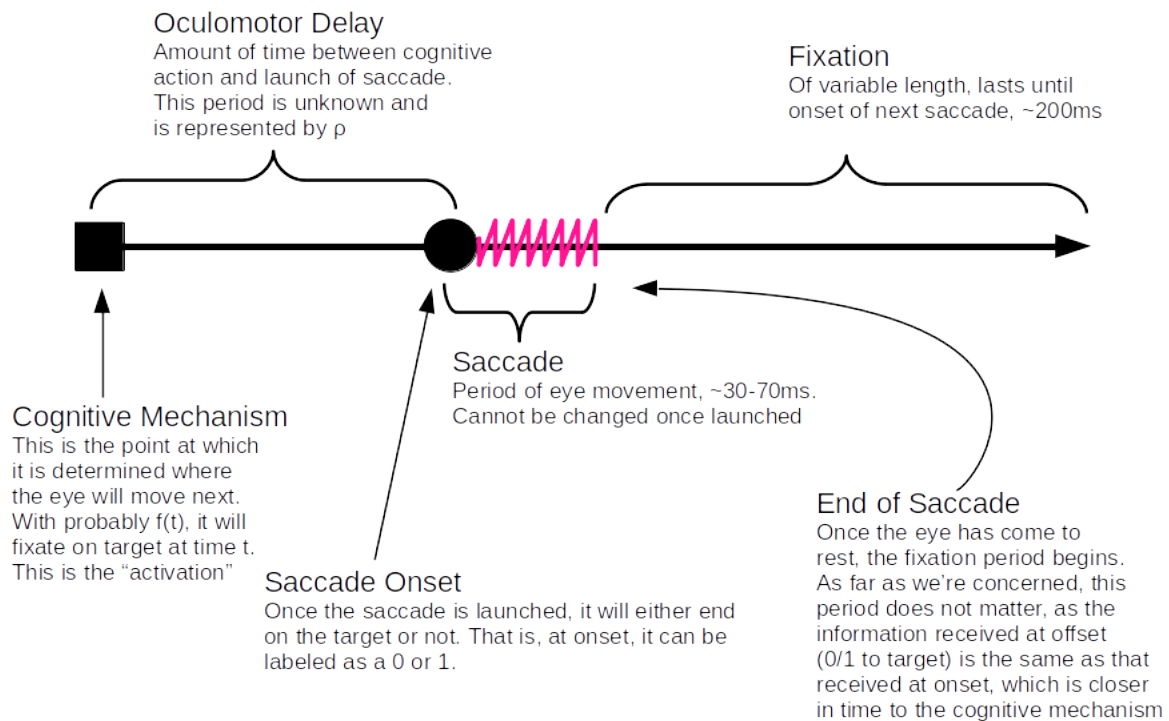


Figure 4: This figure actually doesn't look too bad, but may be better when articulating how saccades measured and why (also includes info on $f(t)$, ρ , etc., so maybe we will present this later around the time of simulation. Mostly here now just to be present

3.2 Activation

What is it, exactly? This would be a good section to introduce notation, specifically that throughout this we will let $f(t)$ be activation in *time*, and in particular $f_\theta(t)$ where

$$f_\theta(t) = \frac{p - b}{1 + \exp\left(\frac{4s}{p-b}(x - t)\right)} + b. \quad (1)$$

Then I can reference Figure 1. Great, references established.

3.3 VWP data

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 Allopenna, a “proportion of fixations” curve is created, aggregating at each discrete time point 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:

$$y_t = \frac{1}{J} \sum z_{jt} \quad (2)$$

where z_{ijt} is an indicator $\{0, 1\}$ in trial j at time t and such that we have an empirical estimate of the activation curve,

$$f_\theta(t) \equiv y_t. \quad (3)$$

In other words, we see here that it is implicitly assumed that the trajectory of the eye follows the trajectory

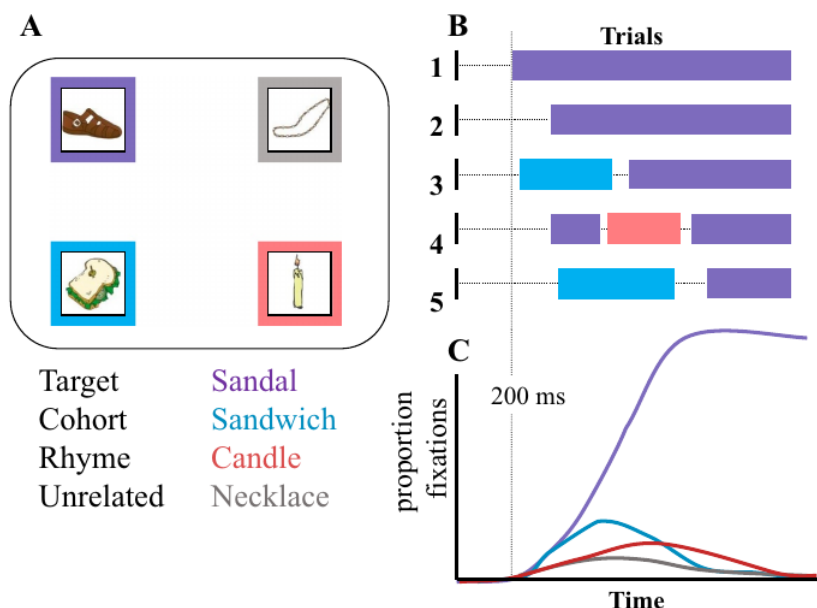


Figure 5: This is screenshotted from Bob’s princess bridge paper. I would like to reconstruct a similar illustration here as it does a great job illustrating the point. *However*, this section as it stands may make more sense elaborated elsewhere, in particular where I give a mathematical treatment to what the “fixation curve” is

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.

4 Where are we going?

Having given due consideration to the state of things as they are, we find ourselves in a time of moral reflection, reexamining the underlying relationship between lexical activation, the mechanism of interest, and the physiological behavior we are able to observe (here, specifically eye-tracking, rather than discussion on other behavioral tasks, i.e., Spivey mouse tracking). This is referred to in the literature as the linking hypothesis. Can elaborate on Magnuson 2019 to whatever degree relevant.

In particular, here we consider a contribution presented by McMurray 2022. From the abstract of his paper: “All theoretical and statistical approaches make the tacit assumption that the time course of fixations is closely related to the underlying activation in the system. However, given the serial nature of fixations and their long refractory 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 have been made. Our intention is to revisit some of the questions raised in this survey and to start towards introducing a constructive path for moving forward. The assumptions made and the general arguments presented can be summarized briefly.

First, we begin with the assumption that there is some generating curve mediating the relationship between activation and saccade generations, and although mechanics are introduced to demonstrate increasingly complex behaviors, these themselves operate independently of the generating function. In this sense, the assumptions here are consistent with those presented by Allopenna 1996 in which word recognition runs parallel with input from visual stimuli. In other words, activation proceeds independently of what objects may have been seen or recognized, and having seen the target object at one instance has no accelerating effect on the rate of activation. Beyond this, there is an accounting for oculomotor delay using a fixed value of 200ms. Finally, there is the introduction of increasingly complex eye mechanics, differentiating in time the duration of the fixations and specifying at what point in time the destination a particular saccade was determined.

Being mostly narrative here, I won't elaborate too much further for now. It suffices to address those points crucial for understanding the direction and purpose of the methodology being proposed. The question addressed is this: in light of the assumptions just described and under increasingly complex conditions, are we able to recover the underlying dynamics of the system in question (activation) given that the "nature of the fixation record [is a] stochastic series of discrete and fairly long last physiologically constrained events?" In short, the answer is no.

McMurray notes that the typical, unspoken assumption implicit in VWP is what he calls the "high-frequency sampling" (HFS) assumption, which states that the underlying activation at some time determines the probability of fixation. This again parallels the assumptions made in Allopenna 1996: "We made the general assumption that the probability of initiating an eye movement to fixate on a target object o at time t is a direct function of the probability that o is the target given the speech input and where the probability of fixating o is determined by the activation level of its lexical entry relative to the activation of the other potential targets." McMurray 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 with the observed data in the context of the HFS assumption relative to other, more complex assumptions. Not sure how much detail is necessary here, but the critical things to note are this: the sources of bias intentionally introduced in the princess bride simulations can be described by two mechanisms: a fixed delay bias, through the introduction of oculomotor mechanics, and a random delay bias introduced by the random duration of fixations and their relationship to when the destination of a saccade movement was generated as well as when it was

observed. Notably, the fixed delay bias resulted in no difficulty in recovering the generating curve under the HFS assumption; after accounting for a horizontal shift, the distribution of bias in the estimated generating parameters was generally symmetric and centered about zero. This was not the case when the duration of fixations had a direct relationship between the timing of the observed behavior.

From this, and what we ultimately argue here, is that the entirety of the observed bias can be partitioned into two distinct components:

1. The first we will call “delay observation bias”. This can be either a random delay, as was implemented in the FBS/FBS+T methods, or a fixed delay, as was observed in all methods, but most notably under HFS and introduced via the imposed oculomotor delay
2. The second source of bias we call the “added observation bias”. This involves the fact that we are “observing” data points, indicated with $\{0,1\}$ at any time t without having observed any behavior associated with the generating curve at that time. This source of bias will be the primary emphasis for our proposal.

We consider first the delayed observation bias. In the simulations presented, this was introduced through both a fixed delay, meant to simulate the effects of oculomotor delay, as well as a random delay, introduced through a mechanism whereby once a fixation is “drawn”, the subject remains fixed on a particular object for the full length of the fixation, with the following fixation’s location determined at the *onset* of the previous fixation. This follows the idea that once a fixation is made, the subject begins immediately preparing to launch their next saccade.

McMurray demonstrated that under HFS with only a fixed delay, the generating curve was able to be recovered without bias. In reality, an oculomotor delay is either truly fixed, in which case recovery is trivial (and especially in the case of comparing generating curves between groups in which both the magnitude and location of observed differences will be preserved under a horizontal shift), or the delay has an aspect of randomness to it, in which case it simply adds to the already random delay that comes from uncertainty in knowing when the decision to launch a saccade is made. As such, we can evaluate the effect of the delayed observation bias by limiting ourselves to testing two cases: one in which there is a fixed delay (here assumed without loss of generality to be zero) and one in which the delay is random. This has the added benefit of freeing ourselves from having to account for any particular assumptions on the source of this delay, only to say that it exists.

The second source of bias introduced is what I call added observation bias and comes singularly from the fact that we do not differentiate between fixations and saccades in the observed data. To illustrate, consider a situation in which there is no delayed observation bias and that the probability that a saccade launched towards that target object at time t is directly determined by the activation of the target at time t , a la Allopenna. When we observe this saccade, s_t , we are directly sampling from the activation curve following some distribution at that point in time,

$$s_t \sim \text{Bin}(f_\theta(t)), \quad (4)$$

where $f_\theta(t)$ is assumed to be the activation curve (elaborated upon in a previous section, the “generating curve” in Bob’s simulation). What, then, to make of the subsequent fixation at time $t + 1$? Under the current method in which the proportion of fixations to the target are computed at each time (which we call the proportion of fixation method), we treat a saccade launched at time t identically with the subsequent fixation at time $t + 1$, up to $t + n$, including the period of time in which there is a necessary refractory period and no new information about the underlying activation could possibly be collected from eye mechanics. An illustration of this bias is given in Figure 6

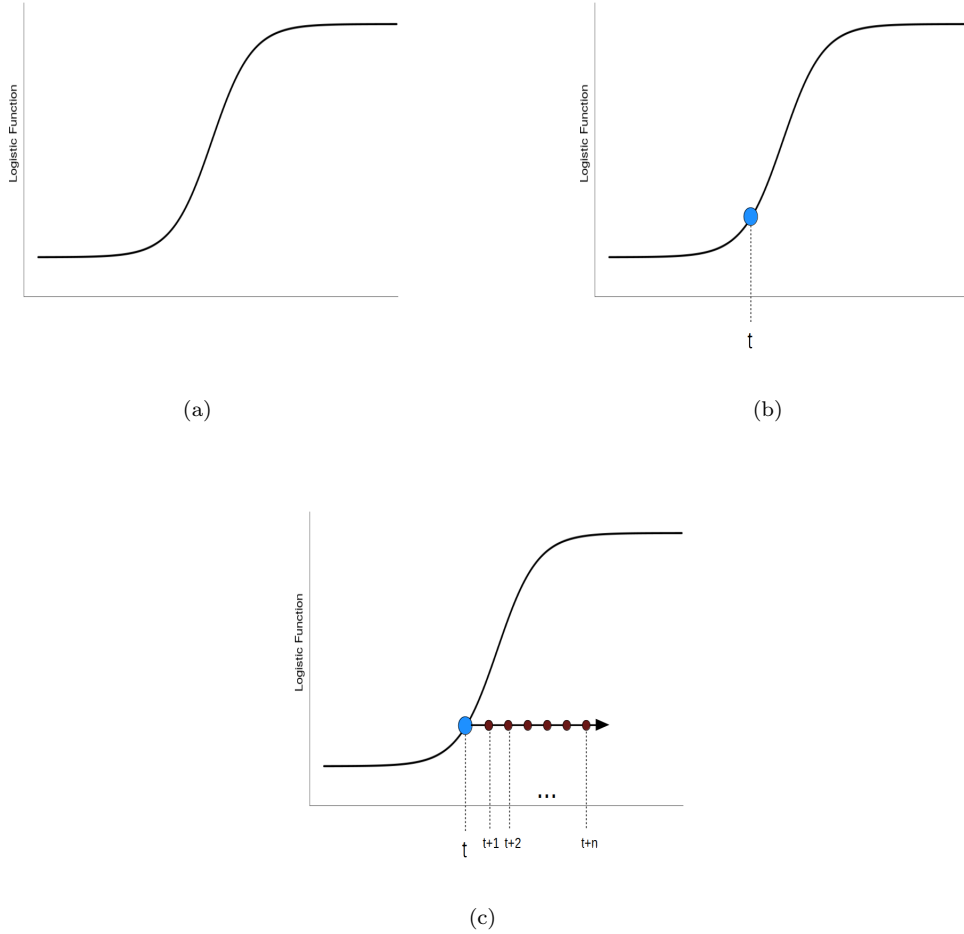


Figure 6: These illustrations can all be made larger (they were made for slides in an image editing program), but they illustrate the main point. **(a.)** here we see an example of a generating logistic function **(b.)** at some time, t , a saccade is launched (in the algorithm, a binomial is drawn with probability $\text{Bin}(f_\theta(t))$) **(c.)** at subsequent times, $t+1, \dots, t+n$, we are recording “observed” data, adding to the proportion of fixations at each time but without having gathered any additional observed data at $f_\theta(t+1), \dots, f_\theta(t+n)$, thus inflating (or in the case of a monotonically increasing function like the logistic, deflating) the true probability.

The consequence of this is that we artificially inflate the *amount* of observed (and also biased) data. And in the particular case of the four parameter logistic function, we artificially *deflate* all of our observations. That is, as our function is monotone, it follows that $f_\theta(t) < f_\theta(t+n)$ for all t and n . As such, a saccade observed at t with some probability $f_\theta(t)$ will also function as an observation at time when the underlying activation is actually $f_\theta(t+n)$, thereby “slowing” the rate of activation. As we will see in the simulations, the result is a delayed crossover parameter and a flatter slope.

While there is no immediate solution to the delayed observation bias, we argue that the added observation bias can be rectified by using *only* observations from saccades in the recovery of our generating curve. A few details on that next.

Saccade Method: Here are a few points to be made in whatever amount of detail. First, we have to rectify the fact that we are now comparing essentially two different curves: one for the proportion of fixations, the other the probability of launching a saccade. Functionally this may be of little importance. Next, we should mention that we can fit this to the same curve (four parameter logistic) using the exact same methods (bdots). Lastly, we can maybe repeat (or move here) a mathematical description of the saccade method, namely what was shown in Equation 5. This is nice because it lends itself to the argument that this is mathematically tractable in that we are clearly specifying the mechanism/distribution. This is less clear in the fixation method where the empirically observed y_t follows no clear distribution. Finally, we should speak to the fact that we are omitting what appears to be “information gathering behavior”. This was addressed in McMurray 2022. I will elaborate more in the discussion, but in short the idea that there is info gathering behavior information in the fixations violates the assumption that activation is running in parallel from visual stimuli. By introducing the saccade method, we are leaving the fixations as an entirely separate component with some potentially interesting avenues to pursue.

5 Simulations

We are going to attempt to isolate the two types of biases identified in the previous section, along with a comparison of the traditional proportion of fixations method with the proposed saccade method.

The first simulation will include no delayed observation bias – that is, a saccade launched at time t will be drawn directly from the generating curve at time t with probability $f_\theta(t)$, where $f_\theta(t)$ represents the generating function which we are ultimately hoping to recover. Following each saccade will be a fixation of random delay, following a gamma distribution with shape and scale parameters empirically determined (Farris-Trimble et al., 2014) (though in reality, any random distribution will do. Notably, the greater the skew the more pronounced the bias). An indicator of the fixation towards the target will be recorded at 4ms intervals. In this scenario, we should expect the saccade method to asymptotically provide an unbiased estimate of the generating curve. For the fixation method, any observed bias will be the direct consequence of the added observation bias.

In the second simulation, we will introduce a delay observation bias similar to that described in McMurray 2022. Following a fixation, a saccade is generated, though in contrast with the first simulation, its probability

of fixating on the target determined at the onset of the *previous* fixation. Here, both the saccade and fixation methods will demonstrate delayed observation bias, while the fixation method will continue to also demonstrate added observation bias.

These simulations differ from the original simulations presented in McMurray 2022 in a few regards. First, we have removed altogether the oculomotor delay, instead keeping all of the delayed observation bias random. This is a consequence of the trivial recovery that comes from horizontally shifting the underlying curve with fixed delay. Additionally, we have collapsed the complexity in generating eye movements into a single mechanic, consistent in that both are contributing to the same, random delay bias (that is, FBS and FBS+T differ in degree rather than kind). Finally, we limit our consideration to only a single generating function, the four-parameter logistic. This is for two reasons. First, we only wish to investigate the aforementioned sources of bias rather than any behavioral characteristics of the curves themselves. Second (and I can elaborate further) given the sensitivity of the fitting algorithm to starting conditions in `gns1` when curve fitting, the double-gauss remains more of a technical challenge to recover. This is a consequence of implementation rather than anything related to the theoretical discussion entertained here.

Similar to the original, an individual subject begins by drawing from an empirically determined distribution a set of parameters for their generating curves. Saccades were launched at random according to the details just outlined with probability determined by the generating curve. In each trial, a record was made of saccades launched, the time at which they launched, and where they had moved. Additionally, an indicator was computed every 4ms with either a 1 or a 0 to indicate if the current fixation was to the target or not, simulating the data generated through eye-tracking software. Fixations were repeated until the sum of fixations in a single trial exceeded 2000ms. Each subject performed 300 trials, and 1000 subjects were generated.

All saccade and fixation data was then fit to the four parameter logistic function with the R package `bdots` (v2) using the `logistic()` function. Given sensitivity to the starting parameters when fitting the curves and to ensure consistency in the fitting algorithm, both groups were given starting parameters `params = c(mini = 0, peak = 1, slope = 0.002, cross = 750)`. The results were not sensitive to the starting parameters. For curves fit to the fixation data, fitted functions with $R^2 < 0.8$ were discarded; for saccade data, fits were excluded if the base parameter estimate exceeded the base parameter or if the slope or crossover estimates were negative. Only subjects who passed both criteria were included. In all, 996 of the original 1000 subjects were kept under fixed delay conditions and 903 under the random delay conditions.

Lastly, each section will present a histogram of the observed bias in the recovery of the generating parameters, where the bias for each subject is computed as the generating parameter minus the recovered parameter. This means that positive bias is the result of underestimating the true parameter while negative

bias is an overestimation. Each section will also present a representative collection of the fitted curves, both using the saccade and fixation methods, against the original, generating curves. Summary statistics on the quality of fits are reserved for the Results section where some more general comments are made.

5.1 Fixed Delay

Will elaborate on plots, analysis of histograms, etc., but not integral for general organization of paper right yet

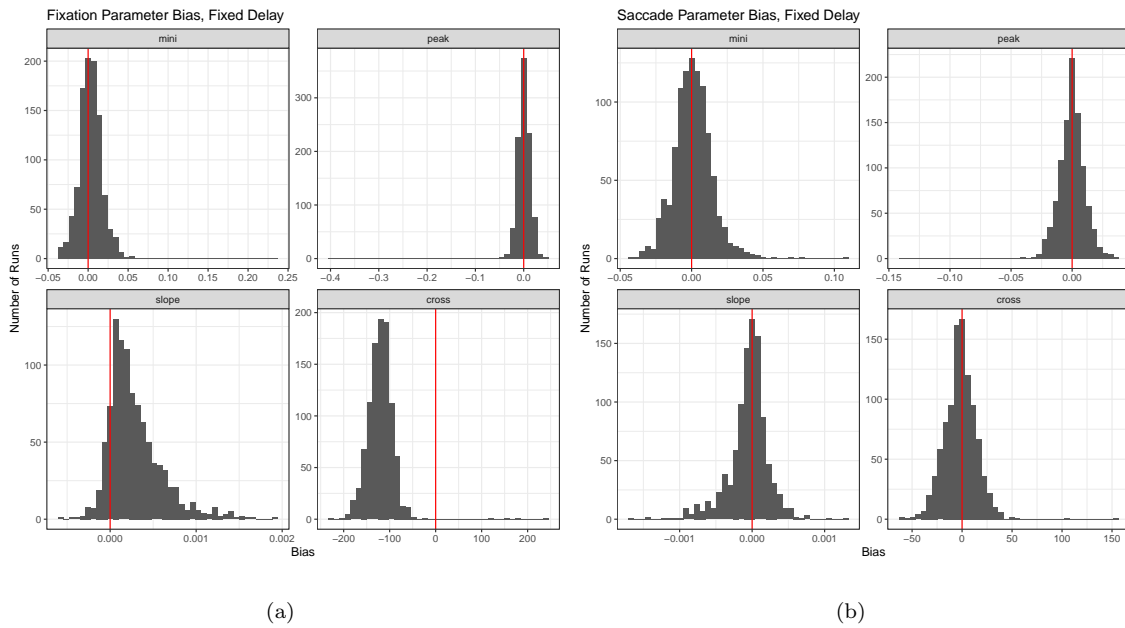


Figure 7: Distribution of parameter bias for fixation and saccade methods under fixed-delay simulation. The bias induced in the fixation method is all a consequence of the added observation bias./ We see evidence that added observation bias has the effect of “pulling” the curve at both ends, resulting in later crossover and less steep curves

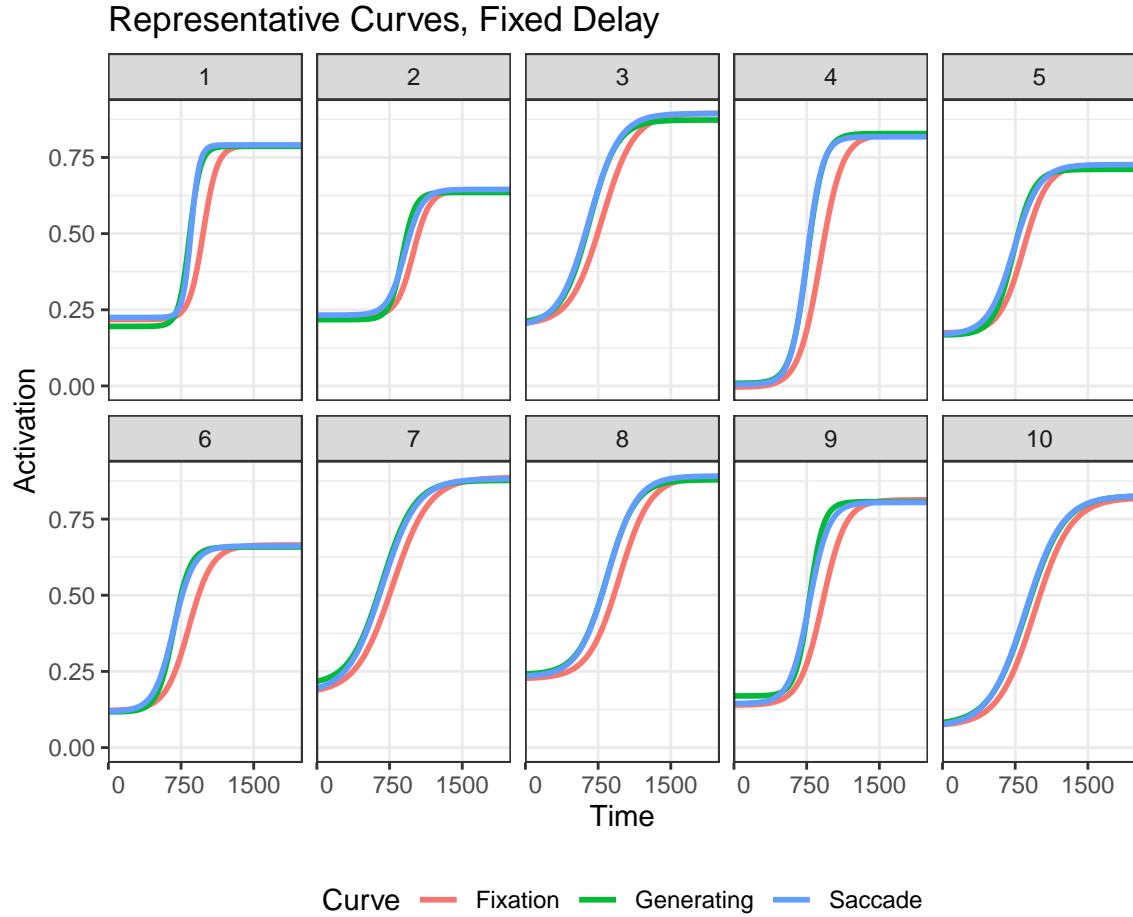


Figure 8: Representative collection of fixed-delay curve, including the generating function, as well as estimated curves from fitting data using fixation and saccade methods

5.2 Random Delay

Will elaborate on plots, analysis of histograms, etc., but not integral for general organization of paper right yet

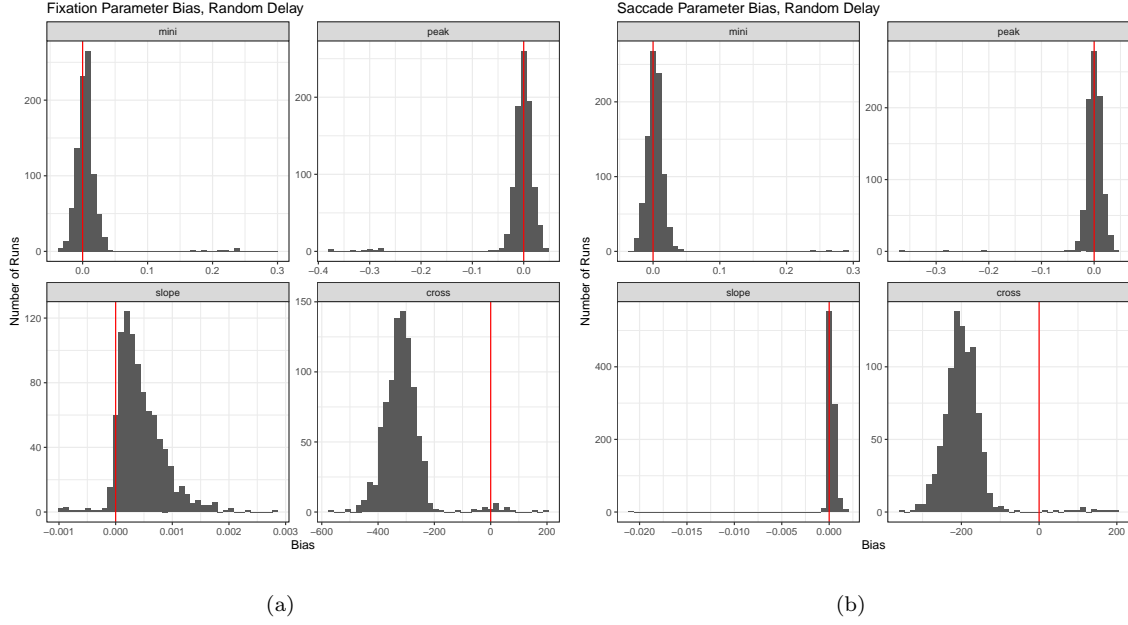


Figure 9: Distribution of parameter bias for fixation and saccade methods under random-delay simulation. The bias induced in the fixation method is all a consequence of the added observation bias AND delay bias, which has consequence of further shifting crossover parameter underestimating slope, but now with saccade too. We see evidence that added observation bias has the effect of “pulling” the curve forward, resulting in later cross over and less steep curves

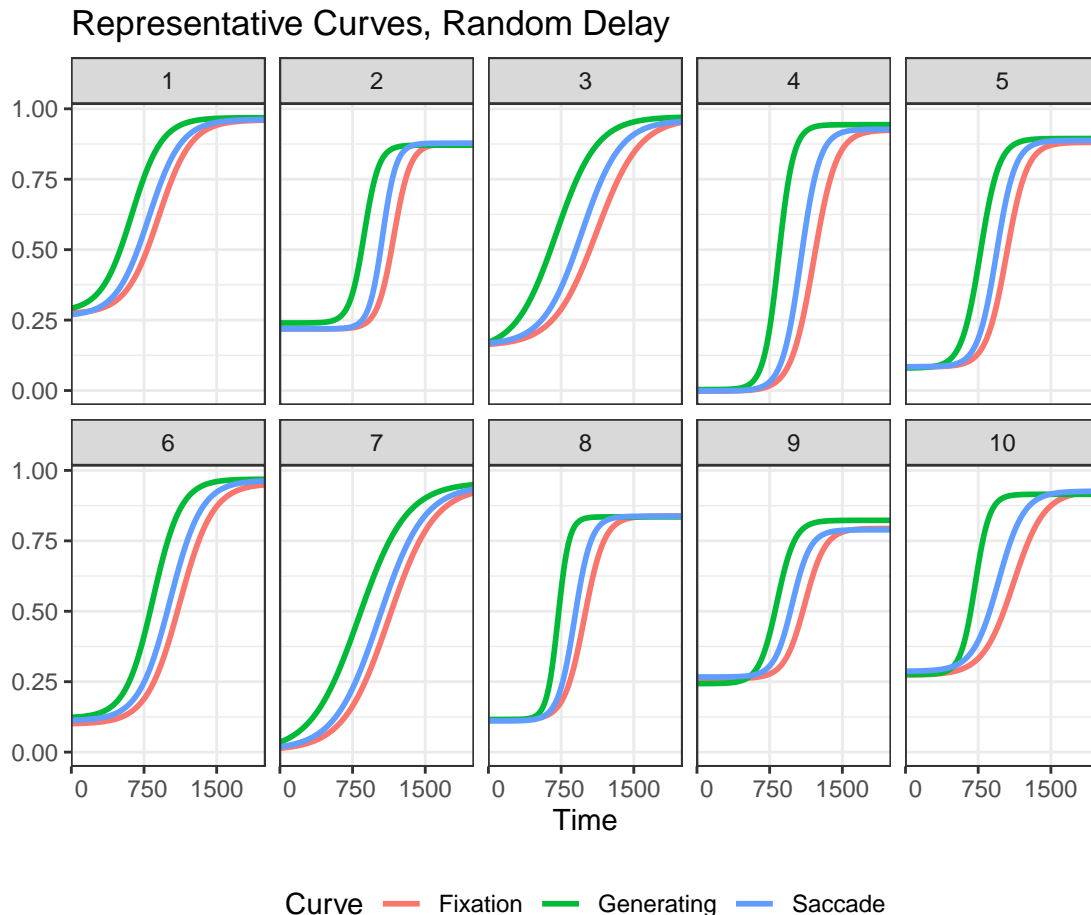


Figure 10: Representative collection of random-delay curve, including the generating function, as well as estimated curves from fitting data using fixation and saccade methods. Note the distortion in shape apparent now in both saccade and fixation curves. Specifically, note how none of the recovered curves are ever steeper than the generating

5.3 Results

Perhaps unsurprisingly, Table 1 demonstrates that (1) situations in which there is no delay between the generating function and observed behavior are easier to recover parameters and (2) the saccade method performed much better in all these cases. This table only includes MISE, I could add R^2 , though the results will functionally be the same.

Curve	Delay	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Fixation	Fixed	1.95	8.18	11.40	13.28	15.98	215.67
Saccade	Fixed	0.01	0.16	0.32	0.52	0.56	78.22
Fixation	Random	20.25	50.95	68.60	73.08	90.92	192.56
Saccade	Random	5.74	21.42	29.29	33.40	40.63	185.79

Table 1: Summary of mean integrated squared error of the fits with their generating curves

5.4 Discussion

This section needs to be tightened and I have said some things elsewhere. Instead, let this be a general collection of thoughts for now.

I would like to speak a little bit more on the concept of “information gathering behavior”. One of the primary benefits of the proportion method is that it indirectly captures the duration of fixations, with longer times being associated with stronger activation. This also 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?). There seems to be a general consensus also that longer fixations correspond to a stronger degree of activation, but a crucially overlooked aspect of this is the implicit assumption that fixation length and activation share a linear relationship. Specifically, insofar as the construction of the fixation curves is considered, a fixation persisting at 20ms after onset (and well within the refraction period) is considered identical to a fixation persisting at 400ms. More likely it seems this would be more of an exponential relationship, with longer fixations offering increasingly more evidence of lexical activation. By separating saccades and fixations at the mathematical level, we are able to construct far more nuanced models (one proposal, for example, might be weighting the saccades by the length of their subsequent fixation, or perhaps constructing a modified activation curve $f_{\theta(t)}(t)$ whereby the parameters themselves can accelerate based on previous information. But this is neither here nor there).

Speaking to the mathematical treatment, there is a wonderful simplicity in letting the saccades themselves follow a specific distribution, namely

$$s_t \sim \text{Bin}(f_{\theta}(t)) \quad (5)$$

or, with random oculomotor delay $\rho(t)$ (which I haven’t really elaborated on as a separate mechanism),

$$s_t \sim \text{Bin}(f_{\theta}(t - \rho(t))) \quad (6)$$

This is in contrast to the fixation method, where the proportion of fixation curves can be described

$$y_t = \frac{1}{J} \sum z_{jt}. \quad (7)$$

Here, is there a clear distribution for what y_t follows? Under independence it may be the sum of binomials, but then what can be said about the relation of y_t to y_{t+1} , given that they may or may not share overlapping fixations from different trials? This is addressed to some degree in Oleson 2017, but this seems more of an ad hoc adjustment to account for this in retrospect. In contrast, the proposed saccade method makes no assumption of trial-level relationship and instead considers all saccades over all trials as binomial samples from the same generating curve in time.

This of course does ignore trial/word/speaker variability, but then perhaps it is time that we shift our language to speaking about a distribution of generating curves for a subject rather than a particular level of activation (note too that this utility is also reflected in the conversation regarding p-values against confidence intervals).

The arguments presented here has hoped to satisfy two goals, agnostic to the linking hypothesis or functions ultimately decided upon. Foremost is the recognition that saccades and fixations are governed by separate mechanisms, and treating them as such allows for fewer assumptions. For example, reconsider again the quote from Allopenna 1996:

“We made the general assumption that the probability of initiating an eye movement or fixate on a target object o at time t is a direct function of the probability that o is the target given the speech input and where the probability of fixating o is determined by the activation level of its lexical entry relative to the activation of the other potential targets.”

Under the saccade method, we omit the entirety of “and where the probability of fixating o is determined by the activation level of its lexical entry relative to the activation of the other potential targets” while still retaining the entirety of the utility in fitting *the same non-linear curves* to less of the data. This decoupling allows the typical time-course utility of the VWP to be used in conjunction with other methods treating aspects of the fixations separately.

Second to this, we have put a name to two important sources of potential bias in recovering generating curves in such a way as to be generalizable beyond the specifics of the assumptions of the simulation (both here and in McMurray 2022). The first, of course, addresses what was just discussed in the decoupling of saccade and fixation data. The utility of the second comes in that it makes no assumptions as to the source of the delayed observation, removing (possibly) unnecessary specifications between oculomotor delay and general mechanics when the goal is to simply recover the generating function. This may be less relevant

when the goal of a study is to specifically address the mechanics of decision making (which itself seems to be difficult to pin down).

In short, what we have hoped to accomplish here is not to drastically change the original assumptions presented in Allopenna (1996) and elaborated upon in Magnuson (2019), but rather to qualify them in statistically sound ways. And really, that is pretty much it. Saccade method is neat, works the same way as the proportion of fixation method, has a more justifiable model while reducing assumptions and allowing room for others.

As a not really conclusion, I am sometimes left to wonder to what degree the proportion of fixation method was a “local minimum” is the pursuit of utilizing eye-tracking data. The proportion of fixations created an ostensible curve, prompting McMurray to establish theoretically grounded non-linear functions to model them. These, in turn, were shown to be suitable functions with which to model saccade data over a period of trials. Had saccades lent themselves so naturally to visualizing as the proportion of fixations, perhaps that is where we may have started.

6 Compare with TRACE

I have a few issues with this section, and as I have fleshed out my reading and understanding of things, my intention with this section has changed. Originally, my hope was to show that non-linear functions fit with empirical saccade data would be a better match to what is predicted by TRACE than what is found using fixation data. This, however, seems to be the wrong thing to do. There is an apparently magnificent number of ways with which to transform TRACE activation data to probabilities of fixation, Neverminding the fact that the saccade curve is a fundamentally *different* concept/mechanism than the proportion of fixations, calling into question the value of a direct comparison (though I’m not sure that this is really much of an issue, as none of these transformations had mechanics uniquely specific to the properties of fixations).

This general idea is related to an observation made earlier by Allopenna and friends,

“It is important to note that although the TRACE simulations provided good fits to the behavioral data, the results should be taken as evidence in support of the class of continuous mapping models, rather than support for particular architectural assumptions made by TRACE.”

Of course, this finds us in a bit of a circular loop – Allopenna suggested that the consistency of evidence was used to support the assumptions of TRACE, and here we are suggesting TRACE as evidence for the saccade method. What seems more appropriate, then, is to demonstrate that there *exist* transformations of TRACE data in which *either* the saccade or fixation methods creates a better fit (as measured by R^2 or MISE). As such, it makes less sense to use TRACE to show which is “better”, but rather to use TRACE to

demonstrate that what is estimated with the saccade method continues to be consistent with the continuous mapping model of lexical activation. Having established theoretical consistency, my argument for the saccade method will rest on what was presented in the previous section, namely the separation of saccades and fixations, and the problems illustrated with the added observation bias.

As it is, I have a few sections here addressing high level concerns. How it should be precisely organized is up for debate, but the work itself should be largely finished.

Finally, I will note here without much other detail – the entirety of this next section rests with the empirical and simulated data from McMurray 2010. For the empirical data, only N/TD subjects were used and for TRACE, only the 14 simulations with the default hyper-parameters from jTRACE. The specifics of data processing (less what I mention in relevant sections here) can be cast to the appendices

6.1 On fitting saccade data

I will go into more detail later on the precise transformations that I did to arrive at the empirical data from the raw data. One key thing to note, however, is that rather than using the separate saccade data in the Access DB, I finished cleaning the fixation data and then transformed this to saccade data based on the start time of subsequent fixations. This helped address ambiguities that resulted from deciding which saccades to be included. For example, if we neglected to include any saccades that began before the onset of the audio signal, the first saccades recorded (generally) had a probability of fixating on the target of about 0.25, resulting in an empirical curve with a base parameter much closer to 0.2. Additionally, this first saccade often occurred some time after onset, leading to virtually no observed data near $t = 0$. This was addressed by artificially setting the first saccade to have occurred at $t = 0$ with its direction set to match that of the matching fixation at the same time. After making this adjustment, the baseline of the saccade curve matched nearly that of the fixation curve, making the saccade curve more closely match the shape of the fixation curve.

Slightly more of an indulgence was how to treat the end of the saccade curves at each trial. Necessarily in all cases, following the last saccade, recorded fixations were constant until the end of the trial, drastically increasing the “added observation” bias and resulting in a fixation curve with a peak much closer to 1.

On one hand, this could perhaps have been dealt with by addressing response times and making the appropriate adjustments. Or, far more simply (and with fewer researcher degrees of freedom), I simply added one last saccade to the end of each trial with its location being that of the last fixation (typically the target object). The rest of the analysis does not depend on this decision in any fashion, and as the results are functionally the same, I elected to use the saccade curve with the inflated data. This most closely

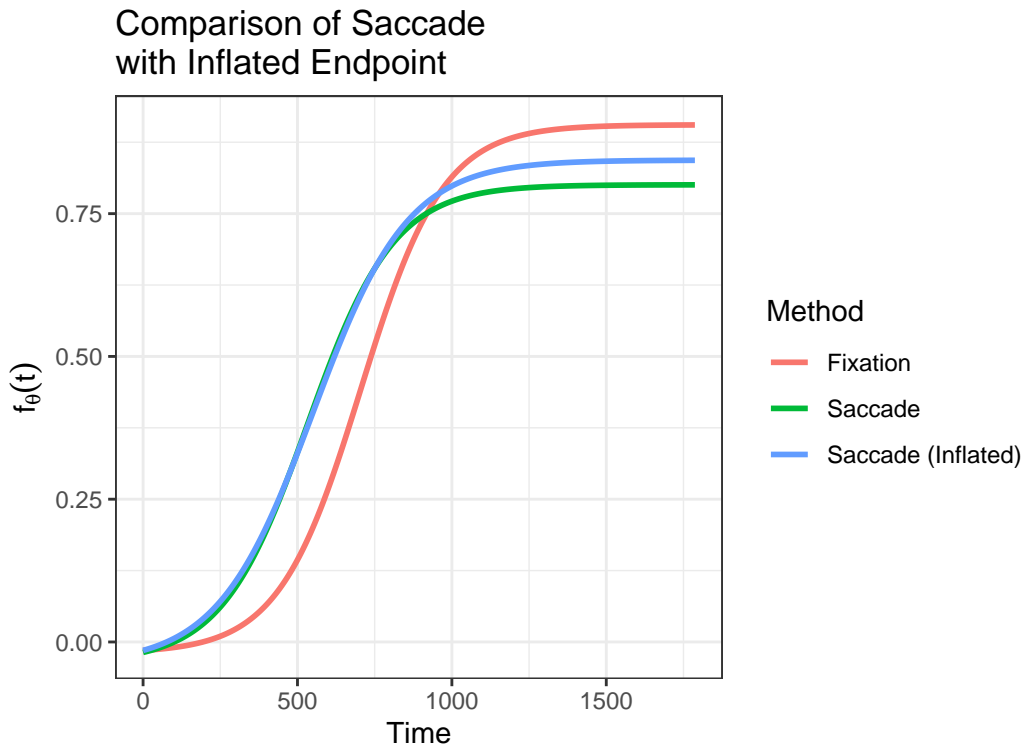


Figure 11: this is what happens when i inflate saccade with additional saccade at last endpoint

matches our expectation of the relation between the fixation and saccade curves and addresses (to some degree) the asymptotic behavior of the saccades which would otherwise be uncollected. A demonstration of these differences is given in Figure 11.

6.2 Transforming TRACE data

My primary concern with the TRACE data is I am seemingly unable to reconcile it visually with what is presented in the 2010 SLI paper. Specifically, I never achieve a baseline near 0. I tried manipulating the temperature of the luce choice rule (LCR) with both constant factors and sigmoidal shapes with differing parameters; I also tried playing with some of the parameters from the scaling factor function.

Referring back to an email we exchanged 12/14/2022, you (you being theBob) gave me a list of adjustments to make to the scaling factor, including swapping the activation and crossover, as well as expanding the exponential term to include the entire denominator. I did this and confirmed that, as you had, the function goes from 0.0002 at $\text{maxact} = -0.2$ to .739 at $\text{maxact} = .55$. The issue, though, is that this is performed *after* luce choice rule implemented. In that situation, the minimum activation observed is 0.25 rather than -0.2. This made me think that perhaps some other permutation of transformations would result in a curve starting closer to 0 and peaking nearer to .75 (for example, scaling the raw TRACE activations). In my

collection of attempts, I never found anything to quite correct for this. It may be a bit much, but I have included plots of the TRACE data related to the target at different points in the transformation process to see how it changes. Maybe something in that will ring at bell. These are included in Figure 6.2.

An interesting aside, though – if we do not make the adjustment to the saccade data where we anchor asymptotic behavior at 0/1, we get a saccade curve bearing less relation to the fixation curve, but with much higher agreement with the set of TRACE curves, in particular with regards to the baseline point and peak. Presumably with some tweaking, it could be made to match even more closely. This phenomenon is illustrated in Figure 12

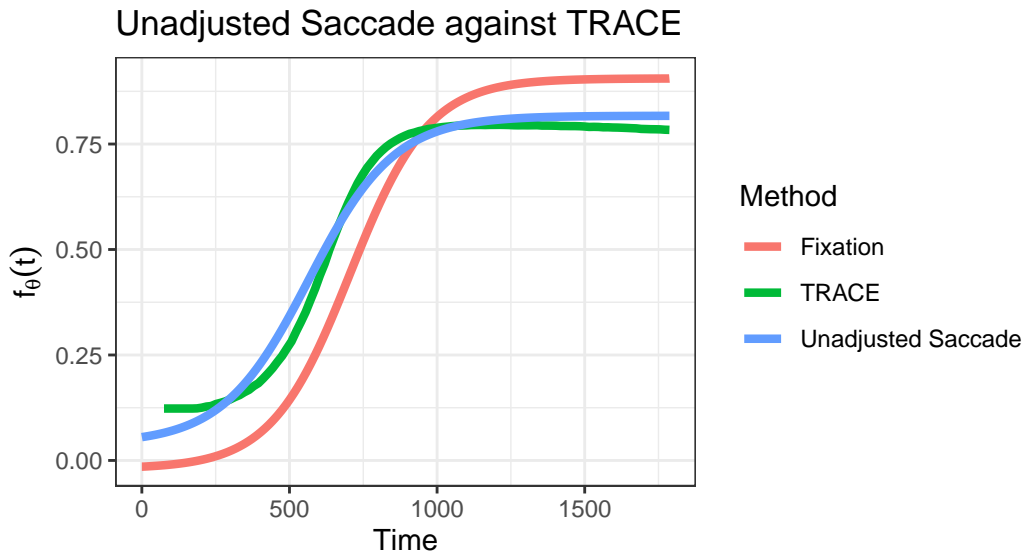


Figure 12: Plot illustrating how the unadjusted saccade method (without anchoring at asymptotes) both matches more closely with the TRACE predictions (particularly near the baseline) while also taking on a far different shape than the fixation curve. This is in contrast to the Princess Bride simulations in which the distortion was minimal and the saccade curve appeared to be more of a horizontal shift

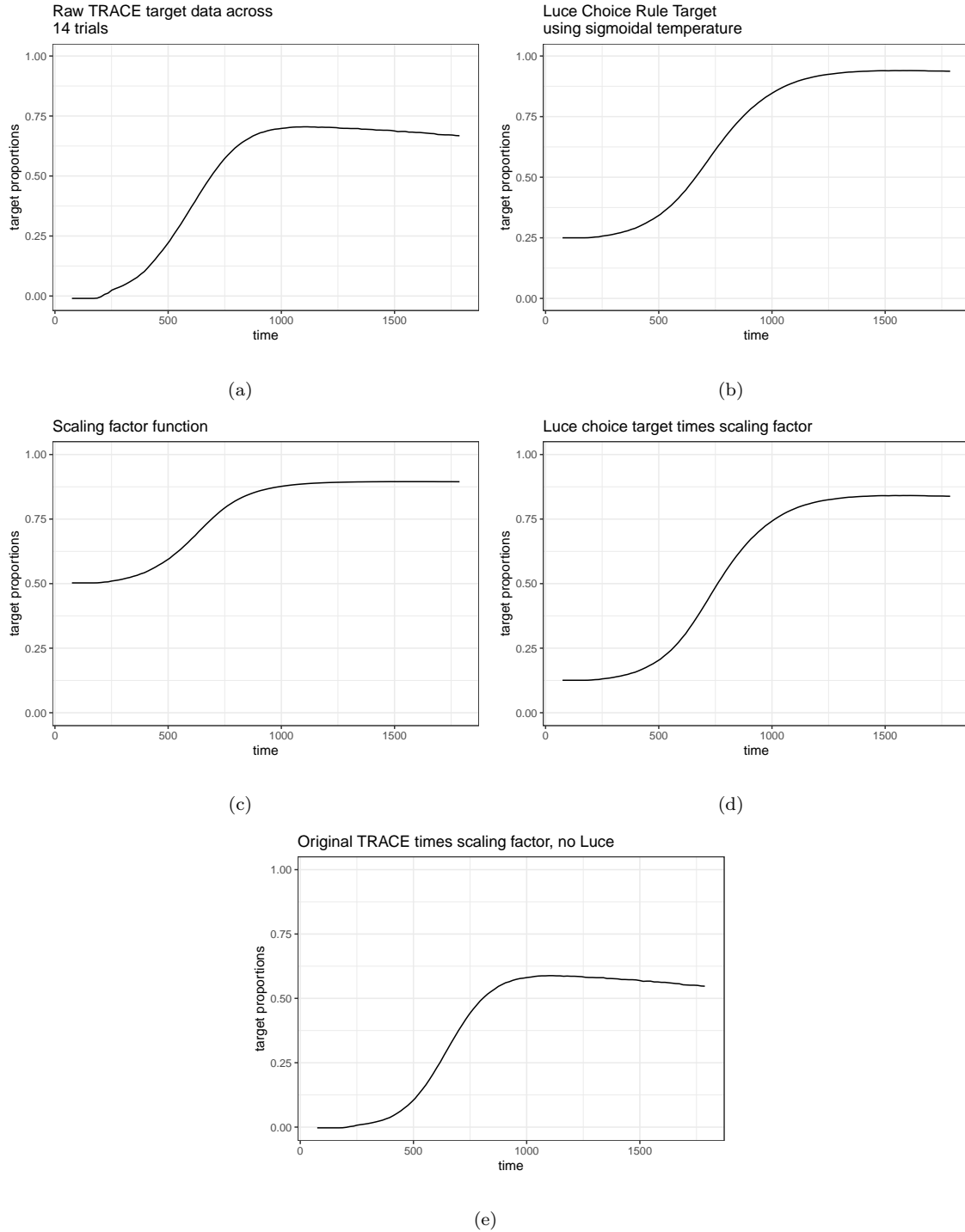


Figure 13: (a) This is simply the raw TRACE data across the 14 simulations with standard parameters. (b) Transformation of TRACE activation using LCR with sigmoidal temperature. (c) Scaling factor function built on max activations *after* performing LCR, using peak/baseline values from target object. (d) TRACE activations following LCR transformation and multiplying by scaling factor. This is what I have been using as model prediction of fixations, though note the baseline value being near 0.15. (e) Perhaps unnecessary, this is simply investigating TRACE activation by the scaling factor but without first conducting LCR. Note that none of these seem to have both the correct baseline and peak values

6.3 Comparisons

Here is where I would suggest the consistency of the models. What I show here is a moderated version of this, namely I show that there are two transformations of TRACE (changing the parameters of the sigmoidal function for Luce choice rule, or the temperature directing the rate at which competitors are weeded out) that each match one or the other of the fixation/saccade curves better. As such, neither is superior in any sense, but both are in the realm of consistency.

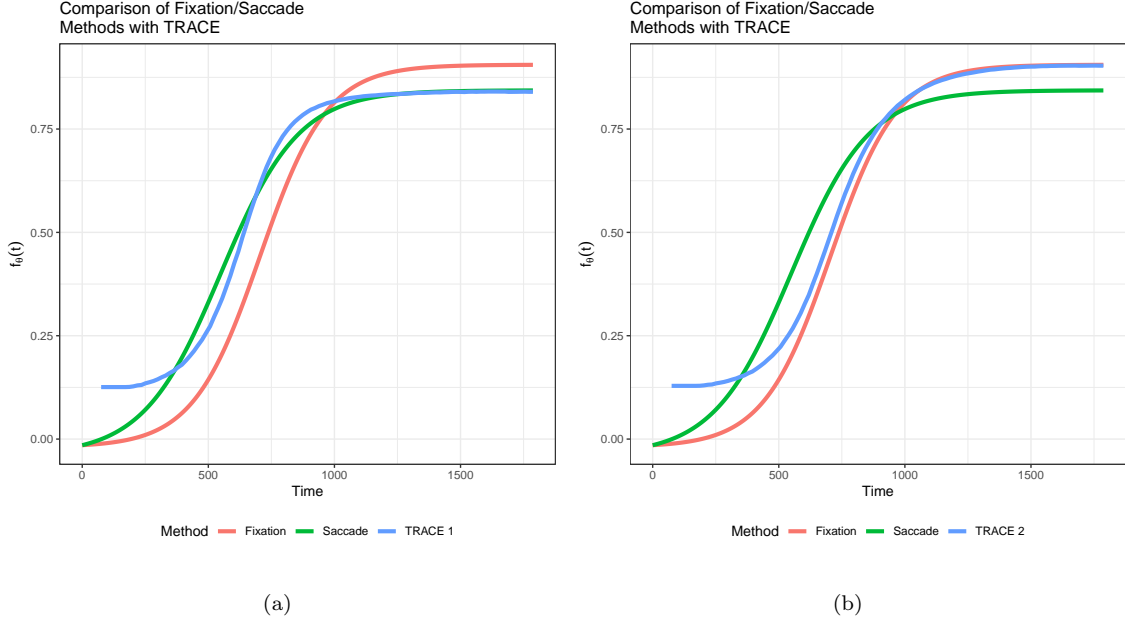


Figure 14: Examples of different temperatures used in LCR and how this effects TRACE activation. In (a), this leads to greater consistency with the saccade curve; in (b), with the fixation curve. This is evidenced also by RMS error values

Presented in Table 2 is a summary of the RMS error of the 1000s simulations using both the saccade and fixation methods against two instantiations of TRACE. As we see and corresponding to (a) in Figure 14 we have better agreement between the saccade method and trace predictions; this relation is flipped for case (b).

	Method	TRACE	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	Fixation	TRACE1	0.1148	0.1743	0.2181	0.2226	0.2583	0.4407
2	Saccade	TRACE1	0.0749	0.1051	0.1396	0.1449	0.1655	0.2933
3	Fixation	TRACE2	0.0991	0.1270	0.1529	0.1606	0.1712	0.3875
4	Saccade	TRACE2	0.0957	0.1404	0.1830	0.1879	0.2275	0.3734

Table 2: Summary of RMS of two transformations of TRACE against saccade and fixation method

As an aside, this also lends itself to the idea of having a *distribution* of curves associated with lexical

activation rather than pursuing point estimation. In some sense, this allows a natural way to account for the observed variability in experimental conditions without having to attempt to model it. Not sure if this is an idea worth elaborating on.

7 Discussion

what have we learned?

Here are really the main takeaways.

1. We are all revisiting question of linking hypothesis
2. In the process of doing so, Bob identified some critical issues, revealing two distinct sources of bias
3. By introducing saccade method, we remove one source of bias and clearly delineate two separate but likely correlated mechanisms
4. This effectively keeps the assumptions from Allopenna and all of the benefits of constructing a function in time for activation, but also allowing room now for fixations to be used separately in a number of ways (length of fixation, latency to look, total fixations, etc.,)
5. Showed that this was still consistent with continuous mapping models by agreement with TRACE

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. Presented in no particular order

Appendix A

Treatment of empirical data from McMurray 2010 to get fixation and saccade curves, along with treatment of TRACE data

Appendix B

I'm not sure if appendix appropriate, but discussion on why double gauss/cohort not considered. This is primarily a consequence of failure to fit adequate models with `bdots`, arising from the fact that `gnls` is highly sensitive to starting parameters. I have demonstrated that they *can* be fit, but successful fits are able to be acquired with a huge range of parameters, bringing into question any validity. As the point of this paper is to demonstrate bias and counter saccade/fixation methods, this seemed an unnecessary addition.

Appendix C

Maybe catch-all for all things OM related. Originally included work showing that fixed delay simply results in horizontal shift, as well as investigation into how the amount of bias is a function both of the length of delay along with the derivative of the generating curve around the delay. Bias near the asymptotes has minimal impact relative to delay near the crossover point.

I think this would be interesting for future research but a bit beyond the scope to detail much here. Could expand on the idea if there was interest as I already have code written up that samples differentially at different time points.