

# What, me saccade?

Last compiled: Tuesday 21<sup>st</sup> February, 2023 at 12:44

## Abstract

In 1995 the Visual World Paradigm (VWP) was introduced blah blah (can I have citations in abstract or should it be more general?)

## 1 Introduction

Spoken words create analog signals that are processed by the brain in real time. That is, as spoken word unfolds, a collection of candidate words are considered until the target word is recognized. The degree to which a particular candidate word is being considered 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.

[start more broadly, there are a number of ways to do this, we use activation]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.

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 → results.

## 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. VWP by Tannenhaus 1995 [?]
2. VWP + TRACE, Allopenna 1996 (trace aspect no longer relevant, just an aside) [?]
3. 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

**Visual World Paradigm** The Visual World Paradigm (VWP) was first introduced in 1995, making the initial link between the mental processes associated with language comprehension and eye movements [?]. A typical experiment in the VWP involves situating a subject in front of a "visual world", commonly a computer screen today, and asking them to identify and select an object corresponding to a spoken word. The initiation of eye movements and subsequent fixations are recorded as this process unfolds, with the location of the participants' eyes serving as a proxy for which words or images are being considered. This association was first demonstrated by comparing how the mean time to initiate an eye movement to the correct object was mediated by the presence of phonological competitors ("candy" and "candle", sharing auditory signal at word onset) and situations containing syntactic ambiguity ("Put the apple on the towel in the box" and "Put the apple *that's* on the towel in the box" in ambiguous scenarios with one or more apples). It is by comparing the trajectory of these mechanics across trials in the presence of auditory or semantic competitors that researchers have used the VWP in their investigation of spoken word recognition.

[“We find that eye movements to objects in the workspace are closely time-locked to referring expressions in the unfolding speech stream, providing a sensitive and nondisruptive measure of spoken language comprehension during continuous speech” [?]]

**Proportion of fixation** 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 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 [?] 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 [?]).

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.

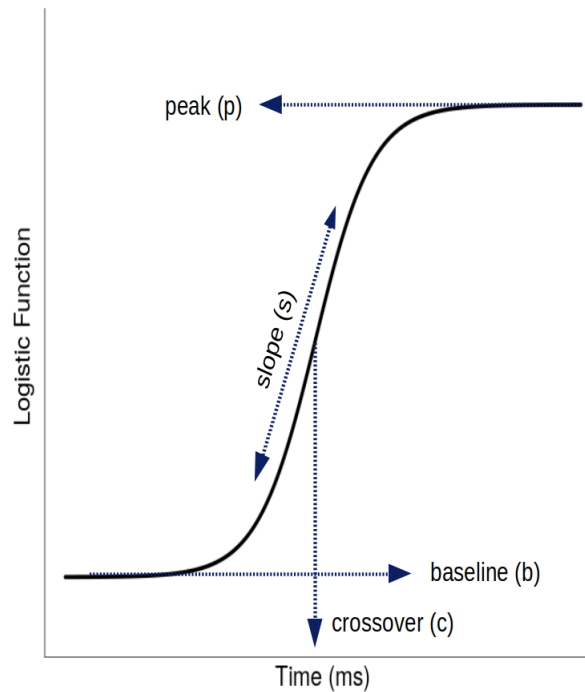


Figure 1: An illustration of the four-parameter logistic and its associated parameters, introduced as a parametric function for fixations to target objects in McMurray 2010. Can describe the parameters in detail, but should also have the formula itself somewhere to be referenced. (Equation 1)

### 3 Where we are now

The following section goes into more detail on the specifics

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

#### 3.1 Anatomy of Eye Mechanics [\[this section needs new name\]](#)

In the context of eye tracking data and word recognition, there are a few mechanics with which we are concerned. The first of these is activation [which i need to learn a little bit more about first]. Even with the immediacy and (fullness? some word they use to describe dense time series here being better than yes/no response), what we observe with any eye movement is not a direct readout of the underlying activation. Rather, there is a period of latency between the decision to launch an eye movement and the physiological response, a period known as oculomotor delay. And finally, there are the physical mechanics of the eye movements themselves, the saccade and the fixation which, together, make up a “look”. We will briefly address each of these in the reverse of the order in which they were introduced.

**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.

Saccades are short, ballistic movements lasting between 20ms-60ms, during which time we are effectively blind. Once in motion, saccades are unable to change trajectory from their intended destination. Following this movement is a period known as a fixation, itself made up of a necessary refraction period (during which time the eye is incapable of movement) followed by a period of voluntary fixation which may include planning time for deciding the destination of the next eye movement; the duration of fixations are typically (some length). It will be convenient to follow previous convention and consider a saccade followed by its adjacent fixation as a single concept called a “look” [?]. We take particular care here to note that the beginning of a look, or “look onset”, starts the instance that a previous look ends or, said another way, the instant an eye movement is launched. A visual description of these is provided in Figure 2.

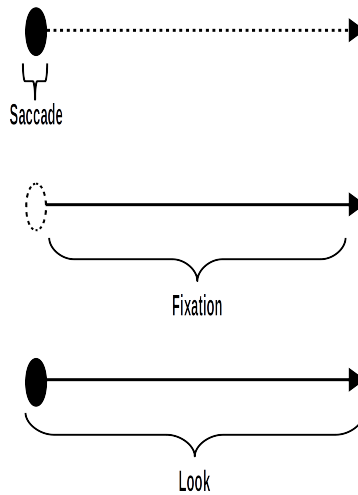


Figure 2: redo this image to match anatomy of look image, also for size

**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 movements in the eye. Between the decision to launch an eye movement (a cognitive mechanism governed by the activation, next section) and the movement itself is a period known as oculomotor delay. It is typically estimated to take around 200ms to plan and launch an eye movement, and this is usually accounted for by subtracting 200ms from any observed behavior. [?]. As oculomotor delay is only roughly estimated to be around 200ms, we suggest that accounting for randomness will be critical in correctly recovering the the cognitive mechanism of interest or at very least in identifying possible sources of bias. How this phenomenon relates to saccades and fixations is demonstrated in Figure 3.

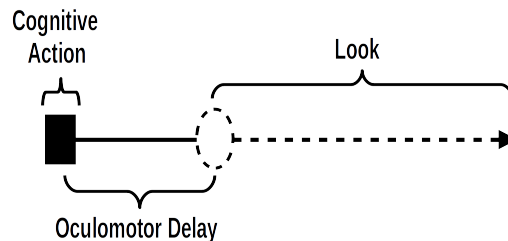


Figure 3: redo this image

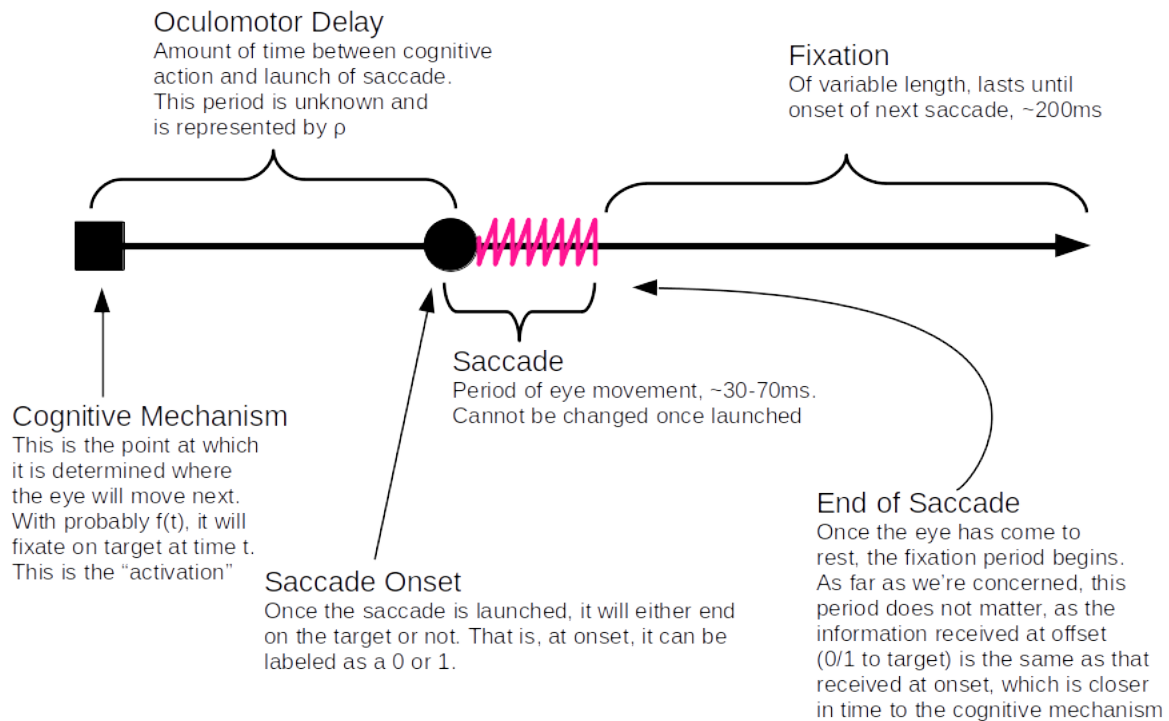


Figure 4: I want to do this figure again but differently. have saccade be two bars matching anatomy of look, include refractory period of fixation, noting that that and saccade are identical, followed by period of time of voluntary fixation (theoretically relevant) followed by next CM

### 3.2 Activation

Here tie in idea of activation, though need to be more concise about what we mean than we are currently. Good source for framework being (McClelland and Rumelhard 1981? Rumelhart and McClelland 81 and 82, and mcclelland/elman 86 with trace). They seem to all mention the “interactive activation framework” which may be worthwhile to elaborate on further. For now, assume that we have adequately stated *what it is*.

While a number of experimental methods are used as real-time indices of lexical access (Spivey mouse trials), we concern ourselves here with the use of eyetracking as it relates to activation as first suggested by [?]. Whereas the initial treatment of eyetracking data made no attempt identify or model subject-specific trends, more recent work has made strides in making subject analysis more tractable. Specifically, we adopt the idea that each participant’s results can be fit to non-linear functions who’s parameters describe clinically relevant properties [?]. We will denote this activation function  $f$  with parameters  $\theta$  as a function in time, giving  $f(t|\theta)$

For example, the four parameter logistic function in Figure 1 is often used to model fixations to the target object in the VWP with functional form

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

Similarly, a six parameter asymmetric Gaussian function,

$$f(t|\theta) = \begin{cases} \exp\left(\frac{(t-\mu)^2}{-2\sigma_1^2}\right) (p - b_1) + b_1 & \text{if } t \leq \mu \\ \exp\left(\frac{(t-\mu)^2}{-2\sigma_2^2}\right) (p - b_2) + b_2 & \text{if } t > \mu \end{cases} \quad (2)$$

(I didn’t make a nice graph/label for this).

While both functions are commonly used in the VWP for modeling eye fixations, for simplicity we will limit the primary focus of our discussion, though ultimately our argument is agnostic to the modeling function used, parametric or otherwise. Discussion related to the asymmetric Gauss is treated in the appendix.

### 3.3 VWP data

We now consider how the aforementioned mechanics relate to the visual world paradigm. 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. All the while, participants are



wearing (generally) a head-mounted eye tracking system recording where on screen they were fixated.

An individual trial of the VWP may be short, lasting anywhere from 1000ms to 2500ms before the correct image is selected. Prior to selecting the correct image, the participant’s 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. A single trial of the VWP typically contains 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 be clear, eye trackers themselves only record  $x$  and  $y$  coordinates of the eye at any given time, and it is only after the fact that “psychophysical” attributes are mapped onto the data (saccades, fixations, blinks, etc.). We adopt the strategy of prior work in discussing eye tracking data in terms of their physiological mapping, as this will be crucial in constructing a physiologically relevant understanding of the problem at hand [?].

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 time point the average of indicators of whether or not 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,  $f(t|\theta)$ . See Figure 5. For any subject  $i = 1, \dots, n$ , across times  $t = 0, \dots, T$  and trials  $j = 1, \dots, J$ , a construction of this curves can be expressed as:

$$y_{it} = \frac{1}{J} \sum z_{ijt} \quad (3)$$

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,

$$f(t|\theta_i) \equiv y_{it}. \quad (4)$$

For our discussions here, we will call this the proportion of fixation method.

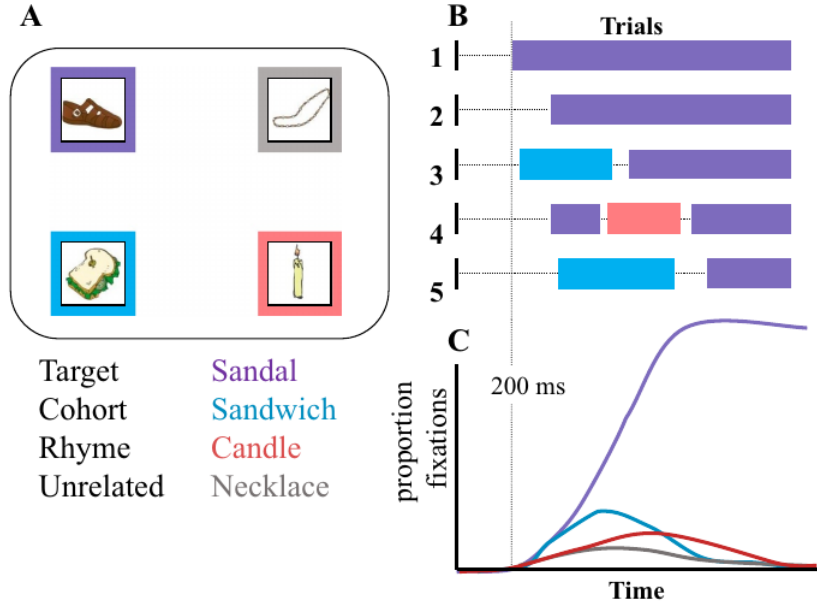


Figure 5: Stole this from Bob (who apparently stole it from richard aslin), plan on making my own

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 are 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. This is referred to in the literature as the linking hypothesis. And while there are a number of competing hypothesis, they each share a collection of implicit assumptions relating what is observed to what is being studied [?].

The simplest version of a linking hypothesis in the context of the VWP is 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 activations of other potential targets (i.e., the other visible objects” [?]. It is from this assumption that we justify the relation in Equation 4. To a degree, this assumption is shared by most linking hypothesis in that the probabilistic nature of the proportions of fixations is assumed to be related in time to the strength of the underlying activation. Primary differences in

linking hypotheses tend to revolve around the particulars of the mechanics involved, including the duration of fixations, eye scanning behavior, the impacts of priming, or the relation between visual processing acting in conjunction with lexical activation.

We consider a particular meta contribution to this debate presented by McMurray in which he probed the relationship between the observed dynamics in the fixations and the underlying dynamics of activation under a variety of assumptions [?]. In short, he showed that curves reconstructed using the standard proportion of fixations analysis in the VWP were poor estimates of the underlying system, with the magnitude of bias increasing on the complexity of the mechanisms involved. Though this made no specific claims as to what the underlying mechanics may be, it did demonstrate the inherent difficulty in relating observable behavior to the underlying cognitive process.

An important contribution made there, however, and one that we adopt here is an explicit definition of the underlying activation function. Given the relation in Equation 4, it is reasonable to assume that the underlying activation of any of the objects with the VWP could be modeled with a nonlinear function  $f(t|\theta)$ . The goal of a VWP analysis, then, is the recovery of this underlying function.

From this assumption we propose an alternative model of the relation between the underlying activation and the observed behavior, with a careful delineation of the psycho-physical components of a look in conjunction with its generating behavior. In particular, we consider the cognitive mechanism associated with initiating an eye movement, which is probabilistically associated with lexical activation, the delay between this and the onset of its associated look, and finally how the different components of the look are related to fundamentally different mechanisms. From this and what we ultimately argue is that observed bias in the recovery of the activation curve under the proportion of fixations method can be partitioned into two distinct components:

[i would like to maybe go into more detail here or have a picture idk]

1. The first source of bias, which is the primary emphasis of my proposal, is what I call the “added observation” bias. This involves the fact that in a standard analysis of VWP data is, the entire duration of a fixation is indicated with a  $\{0, 1\}$  at any time,  $t$ , without having observed any behavior associated with the initiation of an eye movement at that time.
2. The second source of bias is “delayed observation bias”. This bias arises from the fact that an eye movement launched at some time  $t$  was planned at some time prior. This is primarily a consequence of the delay

The first source of bias, the “added observation” bias, arises singularly from the fact that the destination of a look, which is observed at look onset, has a fundamentally *different* generating mechanism than what

determines the duration of a look, never minding such mechanics as the duration of a saccade or the refractory period of a fixation. Nonetheless, a standard analysis of VWP data does not differentiate between the initial onset and the period of subsequent fixation; both are recorded as either 0 or 1 according to its location. A look onset at time  $t$  is probabilistically determined by its lexical activation  $f(t|\theta)$  whereas the period of fixation is governed by a separate mechanism altogether. Treating the subsequent fixation as indistinguishable as the effect of not only “adding” observations to the data, but adding observations that necessarily biased. The result is a distorted estimation of the underlying activation. A depiction of this phenomenon is given in Figure 6.

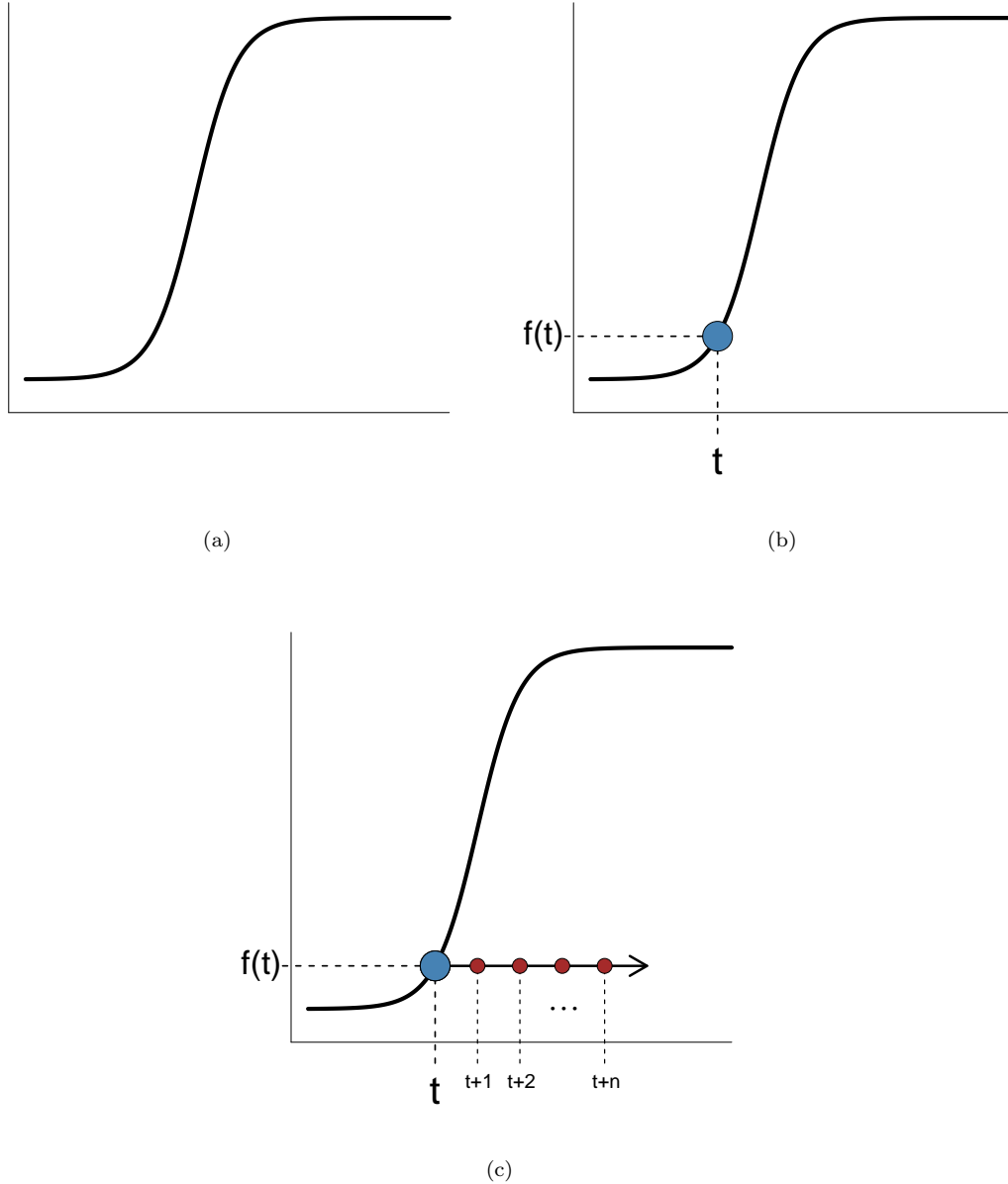


Figure 6: **(a.)** Example of a nonlinear activation curve  $f(t|\theta)$  **(b.)** At some time,  $t$ , a saccade is launched with its destination probabilistically determined by  $f(t|\theta)$  **(c.)** For a look persisting over  $n$  time points,  $t+1, \dots, t+n$ , we are recording “observed” data, adding to the proportion of fixations at each time but without having gathered any additional observed data at  $f(t+1|\theta), \dots, f(t+n|\theta)$ , thus inflating (or in the case of a monotonically increasing function like the logistic, deflating) the true probability.

The second source of bias is the “delayed observation” bias. It is well established in the literature that the time it takes to plan and launch a saccade is around 200ms [?], which is typically accounted for by subtracting 200ms off of the observed data. There are two aspects of this that are worth considering further.

First, if the mean duration of this oculomotor delay is not 200ms, bias will be observed as the difference between the true time and the 200ms adjustment. And although not bias in the technical sense, there has been no accounting for what effect randomness in this delay has on the recovery of the underlying activation. It will be worthwhile in investigating this as the potential magnitude will determine if this delay is worth considering in any more detail in future research.

—

While we present no immediate solution to the effects of randomness in the delayed observation bias, we argue that the added observation bias can be rectified by using *only* the times observed with look onset in the recovery of the underlying dynamics. We call this the “look onset” method, which we explain in more detail.

**Look Onset Method:** The look onset method differ in the proportion of fixation method only in determining which observed data should be considered relevant in the estimation of lexical activation. A particularly compelling argument to made in favor of the look onset method, a corollary of the added observation bias, is that it has a readily defensible mathematical description. A saccade launched at time  $t$  (marking the onset of a look) is assumed to be probabilistically determined by its lexical activation (relative to competitors) at time  $t$ , giving us

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

(it may be that  $l_t$  for look onset is better notation, but my concern is that it doesn’t capture the “onset” nature that we are concerned with and may instead suggest the entire saccade + fixation).

The utility of this is evident when tasked with stating the distribution of  $y_t$  in Equation 3 as it relates to  $f(t|\theta)$ , where given the overlap of fixations within a particular trial, it is unclear what relation  $y_t$  may have to  $y_{t+1}$ .

Two further comments are made about this method here. First, in anticipation of the observation that the look onset method discards relevant information regarding the strength of activation (VOT studies, others from Magnuson review), we acknowledge this and reserve further comment for the discussion. Second given the difference in structure of the observed data, we confirm that the current iteration of **bdots** is capable of fitting nonlinear curves to data both under the proportion of fixation and look onset methods.

## 5 Simulations

Simulations were conducted to replicate the mechanics of a look combined with oculomotor delay, detailed in Figure 7. This section only address Target fixations with a four parameter logistic as given in Equation 1; simulations according to looks to competitors is treated in the appendix. We will begin by describing the process of simulating a single subject.

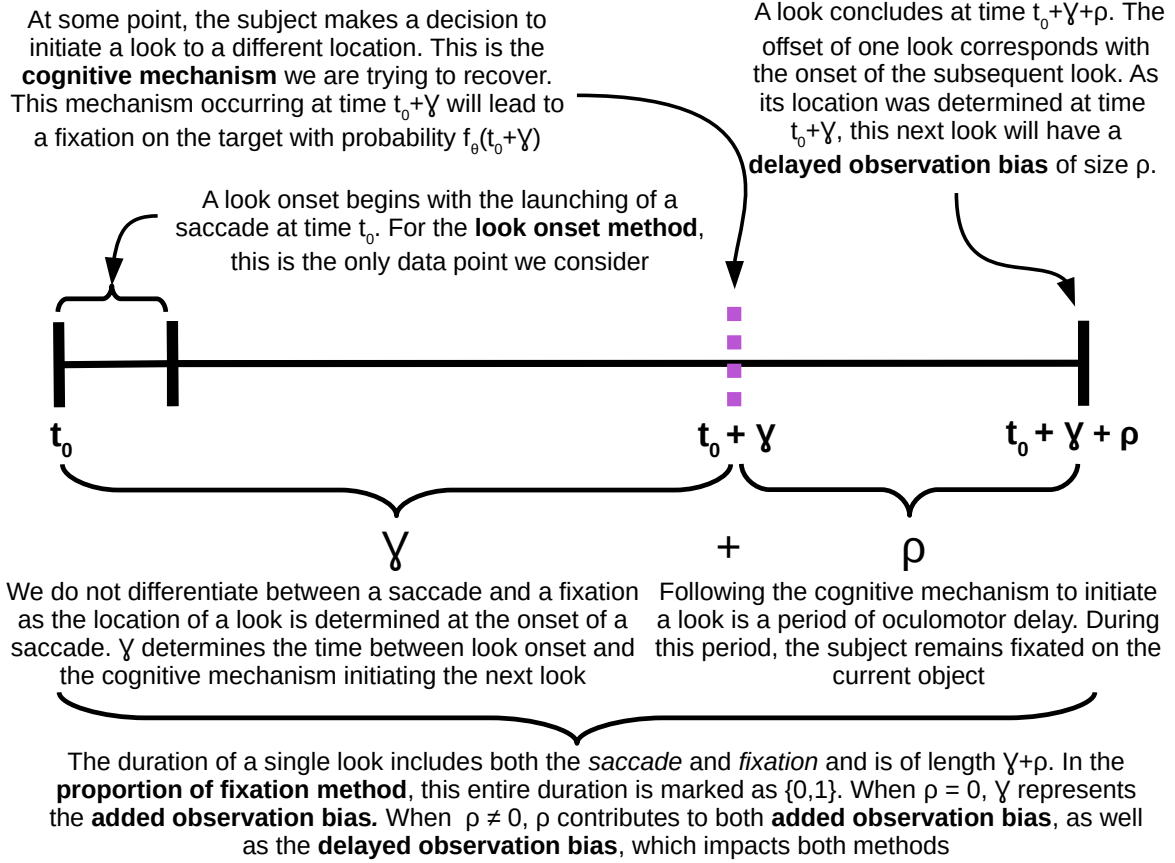


Figure 7: Anatomy of a look – a key thing to discuss somewhere is the OM delay, refractory period, and planning time. The latter two go in  $\gamma$ . Worth noting also that while we do need to be able to control for  $\rho$ , *information* regarding strength of consideration will be in  $\gamma$  - refractory period

First, each subject randomly draws a set of parameters  $\theta_i$  from an empirically determined distribution based on normal hearing participants in the VWP [?] to construct a subject specific generating curve,  $f(t|\theta_i)$ . It is according to this function that the decision to initiate a look at time  $t$  will subsequently direct itself to the Target with probability  $f(t|\theta_i)$ . We then go about simulating trials according to the following method:

at some time  $t_0$ , a subject initiates a look. This look persists for at least a duration of  $\gamma$ , drawn from a gamma distribution with mean and standard deviation independent of time and previous fixations. At time  $t_0 + \gamma$ , the subject determines the location of its next look, with the next look being directed towards the target with probability  $f(t + \gamma|\theta_i)$ . The decision to initiate a look is followed by a period of oculomotor delay,  $\rho$ , during which time the subject remains fixated in the current location. Finally, at time  $t_0 + \gamma + \rho$ , the subject ends the look initiated at  $t_0$  and immediately begins its second look to the location determined at time  $t_0 + \gamma$ . For the look onset method, the only data recorded are the times of a look onset and their location: in this case, at times  $t_0$  and  $t_0 + \gamma + \rho$ . By contrast, the proportion of fixation method records the object of fixation at 4ms intervals for the entire period of length  $\gamma + \rho$ . A single trial begins at  $t = 0$  and continues constructing looks as described until the total duration of looks exceeds 2000ms. Each subject undergoes 300 trials, and 1,000 subjects are included in each simulation.

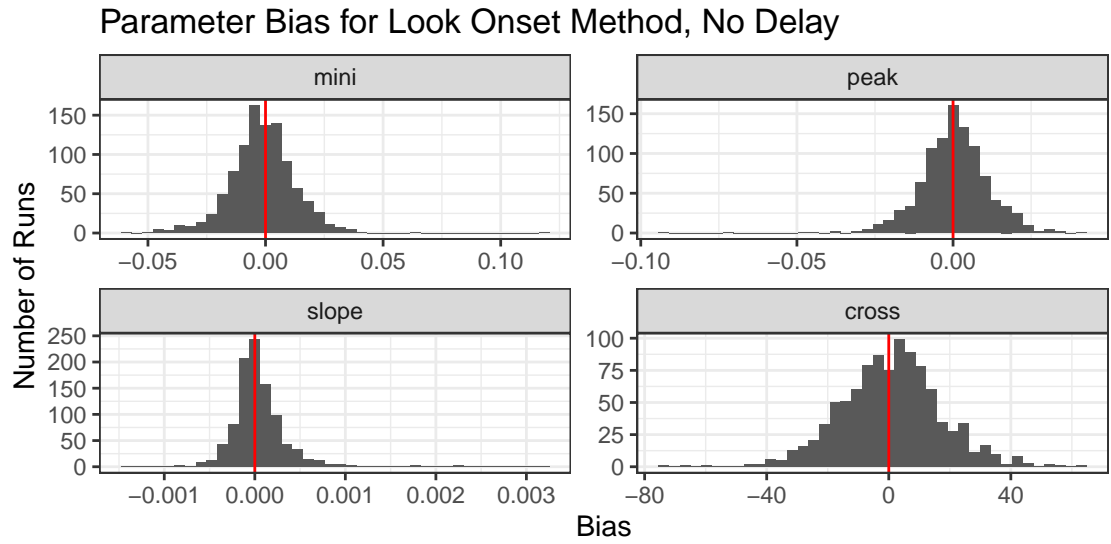
Three total simulations were performed to investigate the biases identified in the previous section, each differing only in the random distribution of the oculomotor delay parameter,  $\rho$ . In the first simulation, we set  $\rho = 0$  to remove any oculomotor delay. In this scenario, a look initiated at time  $t$  by subject  $i$  will be directed towards the target with probability  $f(t|\theta_i)$ . Doing so removes any potential bias from delayed observation and allows us to identify the effects of the added observation bias in isolation. In the remaining simulations we probe the effects of randomness in oculomotor delay, investigating what effect uncertainty may have in our recovery of the generating function. We did this assigning  $\rho$  to follow either a normal or Weibull distribution, each with a mean value of 200ms. As is standard in a VWP analysis, we subtracted 200ms from each observed point prior to fitting the data. Note that a consequence of this is that in these simulations, the bias itself is accurately accounted for by subtracting the correct mean, with the resulting error in the curve fitting process the result of the inherent variability. This does not detract from the argument being made, however, and any true bias in the mean of the oculomotor delay would asymptotically result in a horizontal shift of the observed data according to the direction and magnitude of the bias.

As all of the data could not be individually inspected prior to being included in the analysis, subjects were excluded from consideration if fitted parameters from either the look onset method or the proportion of fixation method resulted in a peak less than the slope, or if the crossover or slope were negative. In the settings in which there was no delay, normally distributed delay, or Weibull distributed delay, 981, 973, and 981 subjects were retained, respectively.

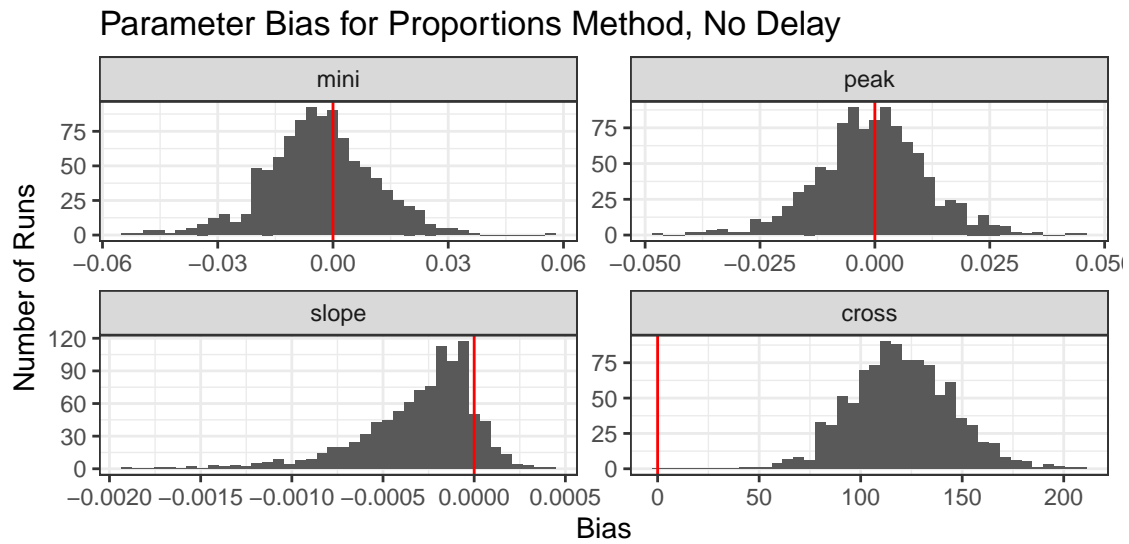
The simulations are performed in R, with the simulation code available on the author's Github page (link?). Simulated data was fit to the four parameter logistic function using **bdots v2.0.0**.



## 5.1 No Delay



(a)



(b)

Figure 8: Parameter bias for no oculomotor delay.

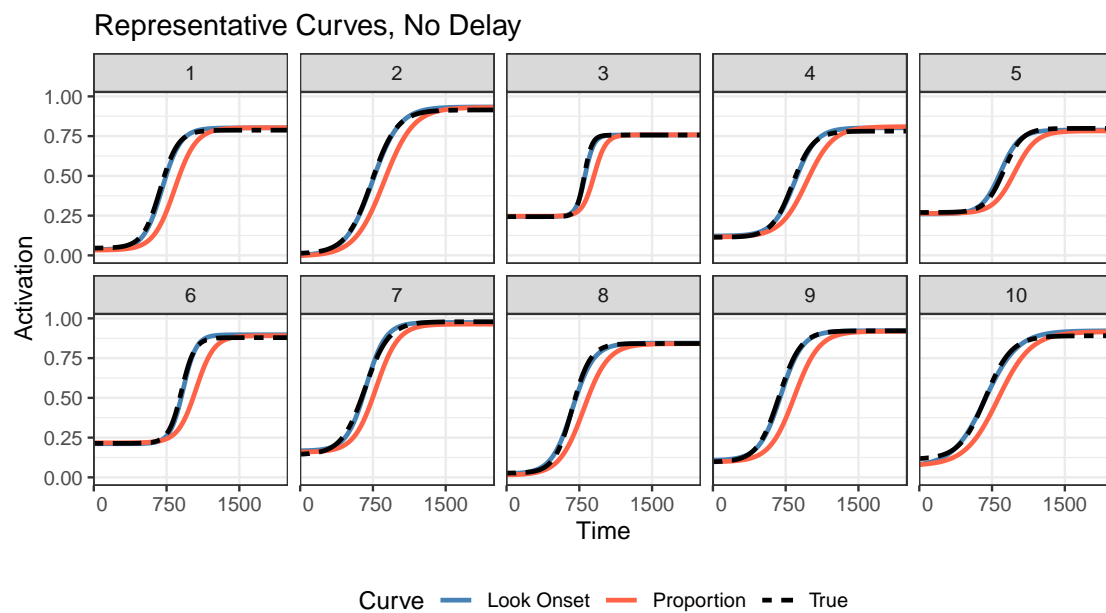
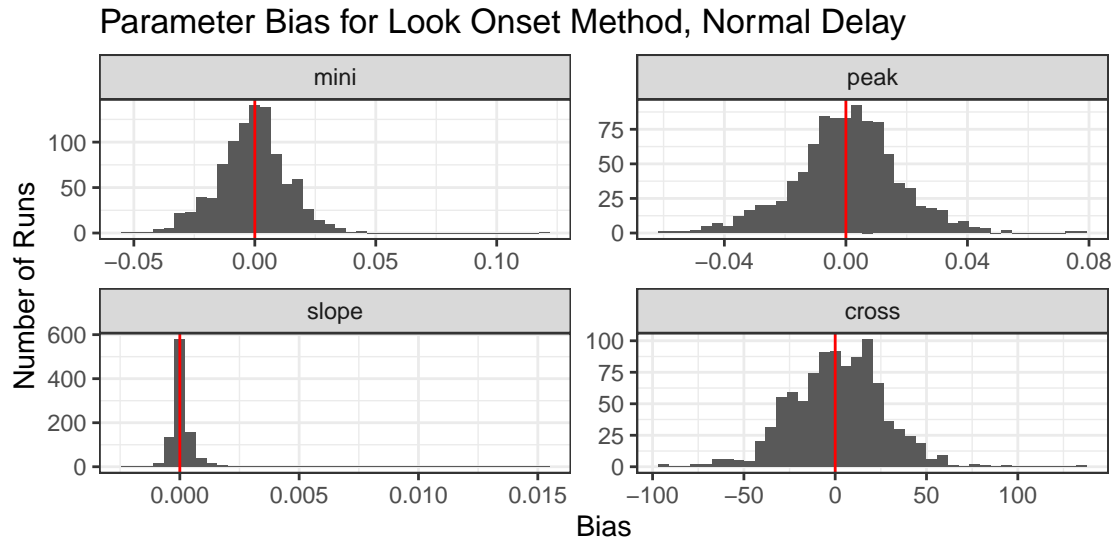
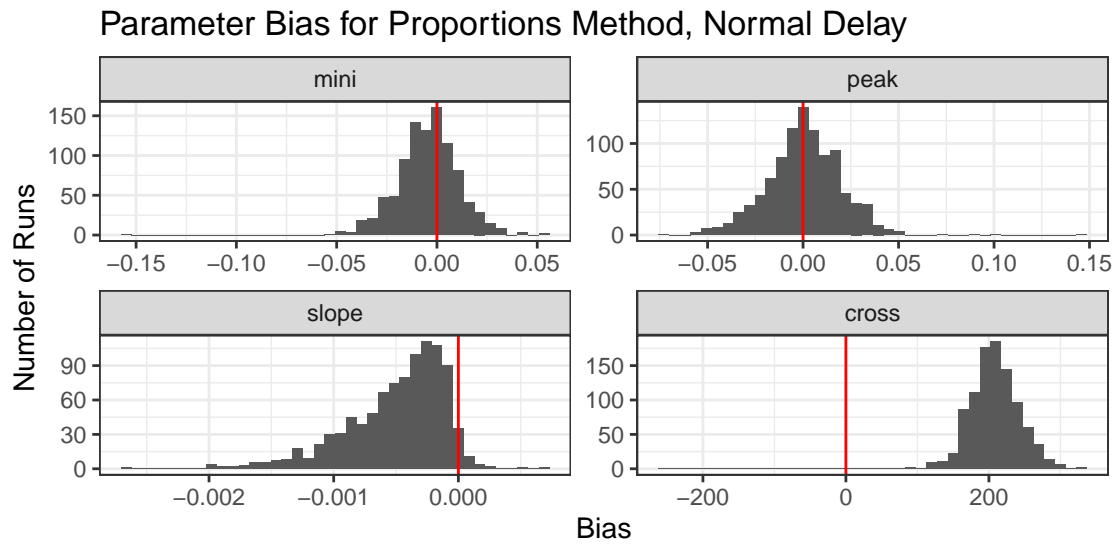


Figure 9: Representative curves for no oculomotor delay

## 5.2 Normal Delay



(a)



(b)

Figure 10: Parameter bias for normal OM delay

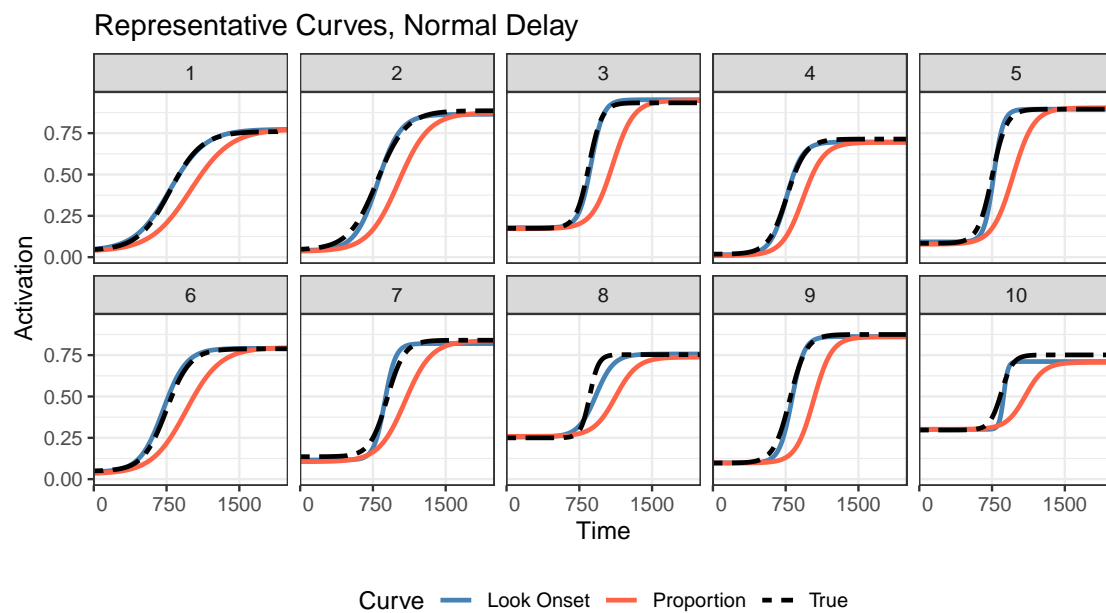
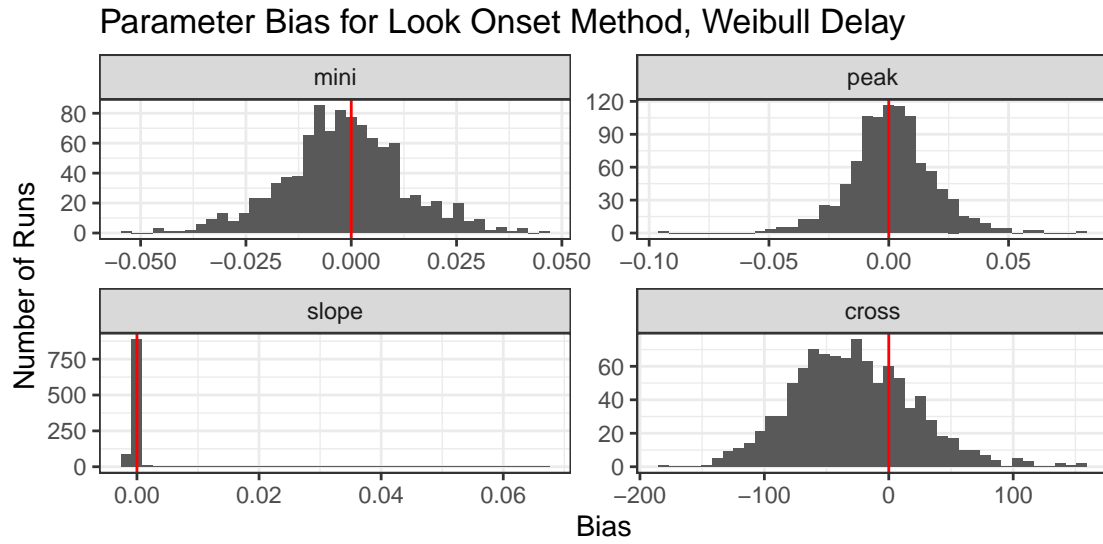
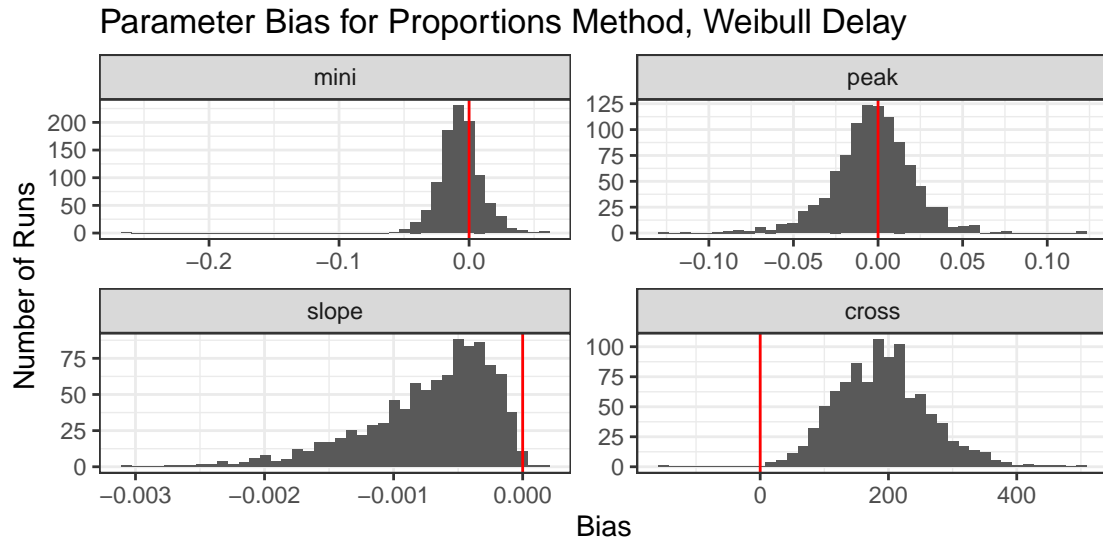


Figure 11: Representative curves for normal oculomotor delay

### 5.3 Weibull Delay



(a)



(b)

Figure 12: Parameter bias for weibull OM delay

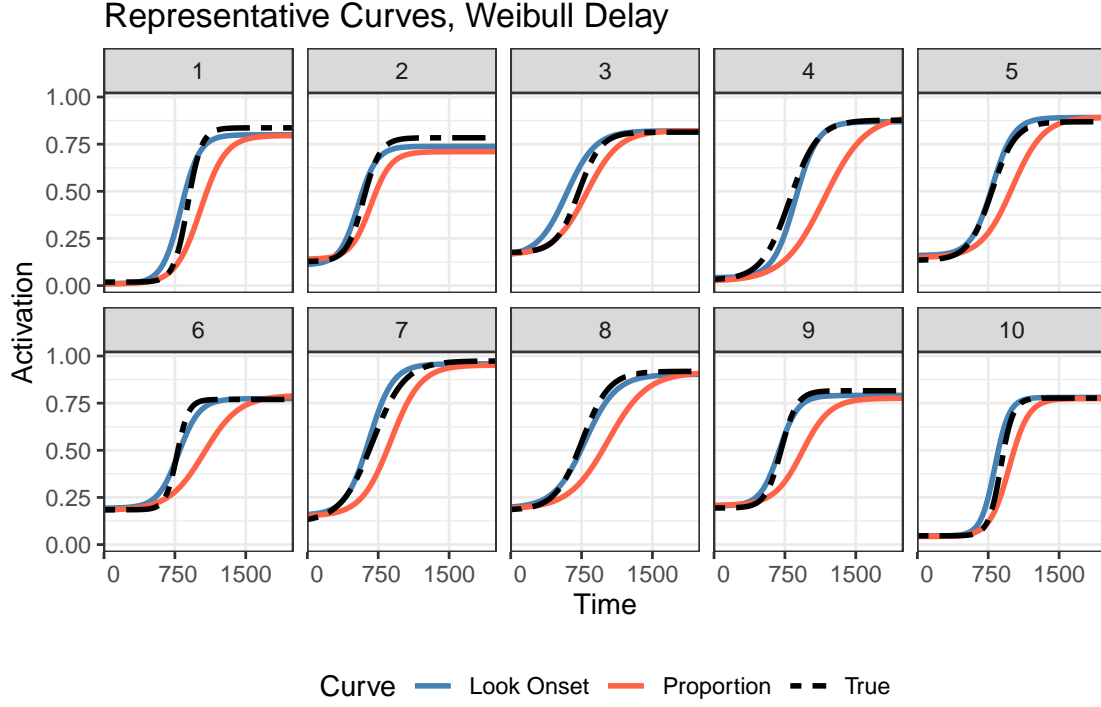


Figure 13: Representative curves for weibull oculomotor delay

## 5.4 Results

Curve	Delay	1st Qu.	Median	3rd Qu.
Look Onset	No Delay	0.17	0.32	0.56
Look Onset	Normal Delay	0.37	0.71	1.24
Look Onset	Weibull Delay	1.05	2.16	4.23
Proportion	No Delay	8.21	11.33	16.01
Proportion	Normal Delay	22.90	30.65	39.37
Proportion	Weibull Delay	15.27	24.75	38.14

Table 1: Summary of MISE across simulations. I don't think I necessarily need (or want) all of those summary stats (min/max specifically, cleaner without)

Unexplored is *where* the delay occurs also important

some concluding remarks about how terrible the proportion of fixation method is

## 6 Discussion

This is now “the” 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

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

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

$$s_t \sim \text{Bin}(f_{\theta}(t - \rho(t))) \quad (7)$$

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

$$y_t = \frac{1}{J} \sum z_{jt}. \quad (8)$$

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.

## 7 Discussion

what have we learned?

No new contributions were added to the linking hypothesis, but introduced novel technique for identifying components of look in VWP and making a standard analysis more consistent with the original

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.,)

## 8 limitations

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

1. linking hypothesis/cognition curve
2. adding parametric form (necessity for saccade method)
3. oculomotor delay, where to discuss
4. Specific results consequence of values chosen and relationship of  $\gamma$  to  $\rho$  (in size, they are already uncorrelated)

## 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 (commented out, input from other tex files)