

# Combining data-oriented and process-oriented approaches to modeling reaction time data

L. ten Bosch<sup>1</sup>, L. Boves<sup>1</sup>, M. Ernestus<sup>1,2</sup>

<sup>1</sup>Radboud University Nijmegen

<sup>2</sup>Max Planck Institute for Psycholinguistics

{l.tenbosch, l.boves}@let.ru.nl, m.ernestus@let.ru.nl

## Abstract

This paper combines two different approaches to modeling reaction time data from lexical decision experiments, viz. a data-oriented statistical analysis by means of a linear mixed effects model, and a process-oriented computational model of human speech comprehension.

The linear mixed effect model is implemented by `lmer` in R. As computational model we apply DIANA, an end-to-end computational model which aims at modeling the cognitive processes underlying speech comprehension. DIANA takes as input the speech signal, and provides as output the orthographic transcription of the stimulus, a word/non-word judgment and the associated reaction time. Previous studies have shown that DIANA shows good results for large-scale lexical decision experiments in Dutch and North-American English.

We investigate whether predictors that appear significant in an `lmer` analysis and processes implemented in DIANA can be related and inform both approaches. Predictors such as ‘previous reaction time’ can be related to a process description; other predictors, such as ‘lexical neighborhood’ are hard-coded in `lmer` and emergent in DIANA. The analysis focuses on the interaction between subject variables and task variables in `lmer`, and the ways in which these interactions can be implemented in DIANA.

**Index Terms:** reaction times, local speed, computational modeling, spoken word recognition, data-modeling versus process modeling.

## 1. Introduction

The advent of ever larger compute power and data sets has intensified the simmering debate between proponents of data modeling or process modeling in all sciences, including psycholinguistics [1]. Behavioral measures such as lexicality judgments and reaction times [2, 3] can be analyzed with linear mixed models, exemplified by `lmer` in [4], but also with process models, such as DIANA [5, 6, 7] that can be seen as an elaboration of older models such as Shortlist-B [8] or SpeM [9]. Regression models ‘explain’ the structure in data sets on the basis of a number of predictors that are selected by the modeler. Process models ‘explain’ –usually small dedicated– data sets by simulating a number of processes that the modeler assumes are going on in the brain. Obviously, data and process modeling reflect deeply diverging approaches. Still, we believe that the two types of model can and must inform each other. Data models can only truly advance our understanding if the predictors can in some way linked to plausible processes. Process models are much easier to develop if we have suggestions about plausible processes that could be derived from data models.

In this paper we present two examples of how mixed effects regression modeling and process modeling can mutually inform each other. The behavioral data under study in this paper come from a priming experiment using auditory stimuli in which subjects provide lexicality judgment and reaction times using prime-target pairs ([10]). The goal of that experiment was to establish under what conditions exemplar effects become visible in the reaction times (e.g., [11, 12]).

The computational model that we use in this paper is DIANA ([5, 6, 7]). DIANA has accurately simulated lexicality judgments and corresponding reaction times for several different lexical decision experiments (Dutch database BALDEY, [13, 5, 6]; the Massive Auditory Lexical Decision (MALD) North American English<sup>1</sup> ([7]). We compare this approach with the results from a regression analysis based on linear fixed effects model (`lmer`) in the R package ([4]).

## 2. DIANA

DIANA aims to simulate participants’ behavior in experiments in spoken word comprehension. The model consists of three components: an Activation Component, a Decision Component, and an Execution Component (cf. Figure 1). Activation and Decision operate in parallel; once the Decision Component has made a decision, the Execution component is initiated. Importantly, like its predecessors SpeM and Shortlist-B DIANA was designed for simulating the processing of stimuli in isolation. Being a computational model of the cognitive processes involved in spoken word comprehension, DIANA does not simulate humanly effects such as waning attention or fatigue.

### Activation Component

The Activation Component uses the signal to compute time-varying activations for all words in the lexicon. The current implementation handle lexicons of 40,000 entries. Entries in the lexicon are phonetically specified as (possibly several parallel) sequences of phone symbols. Each entry is accompanied by a prior probability, derived from the relative frequency of the word in a text corpus. The word activations are determined each 10ms by combining acoustic bottom-up information and top-down information given the word priors:

$$\log P(\text{signal}|\text{word}) + \lambda \log(P(\text{word})) \quad (1)$$

in which the model parameter  $\lambda$  governs the balance between the bottom-up acoustic information (first term) and the top-down linguistic information (second term).

DIANA’s Activation Component performs two types of decoding. A lexical decoding computes the activations over time

<sup>1</sup><http://aphl.artsrn.ualberta.ca/?p=517>

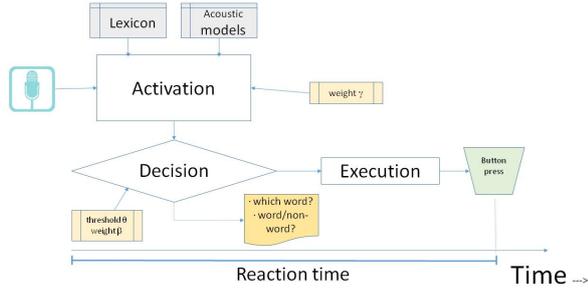


Figure 1: Overview of DIANA. The model consists of three interrelated components: (●1) an Activation Component that takes speech as input; its output is a weighted lattice of hypotheses, evolving over time (●2) a Decision Component, which outputs the recognized word/non-word item and an estimated reaction time (●3) an Execution Component which models the time it takes from the mental decision until the eventual overt action (e.g. pressing a button).

of all lexical entries. Its output is used for simulating word recognition. In the non-lexical decoding the candidates are phone sequences that are phonotactically licensed (constrained) by a probabilistic phone bigram. Both lexical and non-lexical decoding outputs are used for modelling e.g. lexical decision and the detection of unknown novel words.

### Decision Component

DIANA’s Decision Component models two decision mechanisms. First, in DIANA, a decision about the winning word candidate is made at time  $t$  when the activation of the leading candidate exceeds the activations of all competitors with a specified threshold (a model parameter)  $\theta$ . The distance between the leading candidate and the runner-up depends on the word-dependent density of the lexical neighborhood and therefore varies in a non-linear way over time.

If the set of likely candidates is too dense at word offset to make a decision, DIANA adds additional reaction time to simulate the cognitive processes after word offset. The additional RT is modeled based on Hick’s law ([14]), using the entropy  $H(p_1, \dots, p_K)$  involved in the choice of the winning candidate among  $K$  candidates with probabilities  $p_1, \dots, p_K$ :

$$\text{choiceRT} = \beta H(p_1, \dots, p_K) \quad (2)$$

Hick’s law describes the time it takes to process cognitive information in choice reaction experiments, by modeling the depth in a postulated underlying cognitive binary “choice tree”.

### Execution Component

The Execution Component models the process from mental decision to overt behavior (e.g., pressing a button). In the present version of DIANA, this component adds a fixed delay to the time it takes the decision component to identify the word. This delay models the time it takes to execute a planned movement.

## 3. The priming experiment data

The lexical decision experiment that we take as basis for simulation is experiment 1 in [10]. Following [15], they investigated the robustness of exemplar effects in priming experiments. Subjects responded to sequences of stimuli containing

pairs of words with varying degrees of reduction. The decision on the second occurrence of a word (the target) may be facilitated by the first (the prime). An ‘exemplar effect’ takes place when the RT on a target word is shorter when it is preceded by an acoustically similar prime than when it is preceded by an acoustically less similar prime.

The stimuli consisted of an equal number of Dutch words and pseudo-words; all were tri-syllabic infinitives, starting with an unstressed syllable (prefix) with a schwa (e.g., *beschrijven* ‘to describe’, *vertolken* ‘to interpret’). The pseudo-infinitives did not contain phonotactically illegal phoneme sequences. Stimuli were produced by one male (spk A) and one female speaker (spk B). All primes were existing infinitives. Primes and targets could differ in their degree of reduction; they were chosen to have different realizations of the same word. 48 native participants listened to 288 trials in which 34% of the trials formed word repetitions (similar to [15]). Each participant only heard one speaker.

## 4. Simulation 1

Lexical decision involves a potentially large number of partly independent and partly interacting processes. A mixed effects model that predicts the  $\log(\text{RT})$  sequences in such an experiment might look like:

$$\begin{aligned} \log \text{RT} = & P_{stim,1} + \dots + P_{stim,N} + \\ & P_{subj,1} + \dots + P_{subj,M} + \text{linear interactions} \\ & + \text{non-linear terms} + \text{non-linear interactions} \\ & + \text{noise} + P_{\text{previous decisions}} \end{aligned} \quad (3)$$

in which  $P_{\dots}$  denotes a predictor,  $N$  and  $M$  denote the number of stimulus-specific and subject-specific predictors; the last three terms are a shorthand for all interactions, non-linear terms, perceptual and decision noise ([16]), and parameters related to previous decisions (e.g. ‘previous RT’). It is reasonable to assume that processes that play a role early in the decision will be involved in many interactions; processes that come later may have fewer interactions, but it is questionable whether a regression model is at all capable of modeling the effect of the temporal order in which predictors operate.

Robust estimation of all (non)-linear interactions in Eq. 3 would require vastly more data than is available in psycholinguistic experiments. In practice, all non-linear terms are discarded, so that we end up with a linear mixed effects model with subjects and word as random variables. This resulted in an `lmer` model with speaker (male, female), word frequency, the prefix, the prime-target match/mismatch, the duration of the stimulus, and the reaction time given to the prime as fixed effects, and subject and word as random effects [10]. All non-significant effects and random slopes were excluded:

```
modell = lmer(logRT ~ speaker +
logwordfreqcgn + prefix + variant_match
+ logdurms + logRTprime + logRTprev +
(1|subject) + (1 + logRTprev|word))
```

The output of this model (only the fixed factors) is shown in Table 1.

The question now arises which of the predictors in such an `lmer` model might correspond to processes in a model such as DIANA. We distinguish three types of predictors. The first type, which is most clearly related to a cognitive process, is the use of the “previous RT” (`logRTprev`) in the prediction of the current RT. However, DIANA processes all stimuli independently. Still, below it will appear that there is an interesting link.

Table 1: Results of the `lmer` model `model1` (cf. [10], table 4).

Predictor	$\beta$	$t$	$p <$
prefix (ver-)	0.06	5.9	.0001
speaker (speaker A)	0.13	4.9	.0001
word frequency	-0.01	-2.3	.001
target duration	0.39	14.2	.0001
RT prime	0.16	8.5	.0001
RT previous trial	0.21	10.2	.0001
variant match (mismatch)	0.02	3.5	.0001

Table 2: Results of an `lmer` model `model2`, with the difference between (log) RTs on prime and target as dependent variable.

Predictor	$\beta$	$t$	$p <$
(Intercept)	-0.077000	-7.563	.0001
speaker (spk A)	0.082976	7.492	.0001
variant match (mismatch)	0.023581	3.016	.001
log duration	0.321425	8.358	.0001
logRT previous trial	0.120292	4.369	.0001

The second type comprises predictors related to a meta description of the stimuli. An example in `model1` is "prefix". In regression modeling, such parameters undoubtedly matter and have a significant role in the simulation of RTs. Evidently, in DIANA, these parameters cannot be specified as input.

A third group of predictors covers linguistic knowledge that is input in regression models and that is represented internally in DIANA. Word frequency is an example. In `lmer`, word frequency is used as a predictor; in DIANA the relative word frequency serves as a prior probability during the search ([17]). The actual effect of word frequency on DIANA's RT depends also on the neighborhood density, which in turn depends on the composition of the lexicon. In an `lmer` model neighborhood density must be specified as just another predictor.

The `lmer` model `model1` shows that mismatch between prime and target leads to significantly longer RTs on the target than in the match condition. In order to model this 'exemplar' effect, DIANA should be able to increase the prior probability of the full or reduced version of words as *targets* with the appropriate difference. Exactly how that should be done is not obvious. The `lmer` model `model2`, in which the *difference* between RTs on the prime and target is the dependent variable, shows why adapting the prior is complex:

```
model2 = lmer(difflogRT ~ speaker +
variant_match + logdurms + logRTprev +
(1|subject) + (1+logRTprev|word))
```

The output of `model2`, summarized in Table 2, shows that a mismatch between prime and target (reduced versus unreduced, or vice versa) leads to a significant difference of 0.024 in  $\log(\text{RT})$ . Therefore, the wrapper around DIANA should be able to adapt priors in the lexicon on the basis of the pronunciation *variant* that was recognized, not only on the basis of having the word type encountered previously. We simulated the RTs with DIANA, once with fixed priors in the lexicon, and once with priors that we increased as function of the prime-target mismatch. With fixed priors the correlation between the RTs predicted by DIANA and the average RTs of the participants is 0.34; with adapted priors the correlation is raised to 0.38. This is a promising result, given the fact that the average correlation between RT

sequences of pairs of participants is only 0.085. In this case, the improvement in DIANA as a process model is motivated by the output of data-oriented regression modeling.

## 5. Simulation 2

In the second simulation we study the reverse, process modeling can inform regression approaches. It has been observed that the residual error in sequences of RTs in mixed effects models may contain a substantial autocorrelation [18]. That may even be the case if the model has the RT to the previous stimulus (RT previous), which is supposed to capture what is known as 'local speed effects' [19] in RT sequences, as one of the predictors. In developing DIANA maximizing the correlation between simulated RTs and average RTs of participants was an important criterion. Since DIANA processes stimuli one by one and ignores the previous result, the comparison between RT sequences simulated by DIANA and the observed human RT sequences only makes sense if the impact of the different ordering in which participants heard the stimuli is accounted for. This is done by detrending the human RT sequences by applying an autoregressive (AR) filter on the observed RT sequences. This filter generalizes the 'previous RT' to a weighted RT (denoted  $maRT = maRT[1, 2, \dots, N]$  where  $N$  denotes the number of reaction time measurements in one session), which is based on all previous RTs with a weighting factor that is exponentially decaying with recency:

$$\begin{aligned} maRT[1] &= RT[1] \\ maRT[i] &= \alpha RT[i-1] + (1-\alpha)maRT[i-1] \quad (4) \\ &\text{for } i > 1, \end{aligned}$$

with  $0 < \alpha \leq 1$ . The conventional 'previous RT' is obtained by setting  $\alpha = 1$ . After computing the  $maRT$  sequence, the difference  $RT[i] - maRT[i]$  is taken as the current detrended RT. This detrending removes both the long-term effects on the observed human RT sequence (health condition, age, gender, handedness, general cognitive abilities, etc. [16]) as well as the mid-term effects (attention fluctuation, strategy changes, learning effects, fatigue), leaving the stimulus-related aspects of interest (lexical status, morphological complexity, the density of its lexical neighborhood, frequency effects, indexical effects) (e.g. [20, 21, 22, 23, 24, 25, 26] and references therein). From the literature, it is known that RTs of the previous 5 to 10 stimuli may influence the current RT [26], which corresponds to a value of  $\alpha$  between 0.1 and 0.2.

Pairs of participants often show extremely low correlations in their RT sequences ( $< 0.15$ , [6, 7]). The same is true in the data under analysis here. This is caused by the influence of a multitude of participant dependent long-term and mid-term effects on the RT given to the current stimulus ([6]). It appears that the between-participant correlation can be substantially increased by detrending RT sequences with an AR filter with  $\alpha \approx 0.1$ .

In this simulation, we first investigated the impact of the detrending parameter  $\alpha$  on the average correlation between the observed RT sequences across all pairs of subjects, as obtained in [10]. The results are given in Figure 2. The horizontal axis displays the range of  $\alpha$ ; the vertical axis shows the average correlation. The rightmost point  $\alpha = 1$  coincides with the choice in the `lmer` model in [10], presented in the previous section. The figure shows that, in general, the average correlation increases with decreasing value of  $\alpha$ . This holds for RT data in which participants only heard the male or female speaker

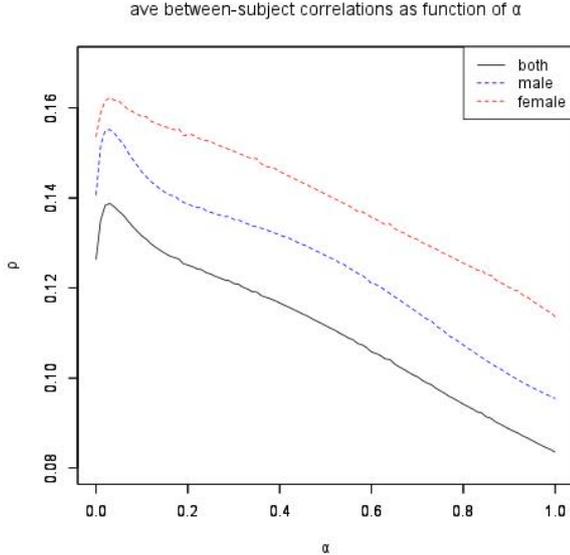


Figure 2: The average correlation between the RT sequences across all pairs of subjects, as a function of the detrending parameter  $\alpha$ .

(dashed lines), as well for the combined model in which data from both male and female speaker are combined (solid line). The optimum is reached for  $\alpha \approx 0.05$ , showing that in this lexical decision experiment the reaction times of about 20 previous trials play a role in the current RT. The same trend has now been observed for RT data in many different lexical decision experiments ([5, 7]). It clearly shows that the way in which the modeling of the local speed is dealt with may have a serious impact on the match between the RT sequences simulated by a model and obtained from participants.

Second, we studied the impact of  $\alpha$  on the significance of the predictor match/mismatch. Figure 3 shows the value of the  $t$  statistic as provided by `lmer` for the predictor match/mismatch, on which the existence of the exemplar effect in this priming experiment ([10]) hinges. The  $t$  statistic is displayed along the vertical axis. The dashed lines show the  $t$  statistic as function of the detrending parameter  $\alpha$ ; the thin horizontal lines indicate the constant values independent of  $\alpha$  as in the `lmer` model from [10]. This figure clearly shows the dependency of the significance of the match/mismatch condition in prediction of RT as a function of  $\alpha$ . One observes that the significance is also gender dependent. On the combined set, there remains a clear significance. The significance is larger than reported in [10] if  $\alpha$  is chosen smaller than 0.2, because this value of  $\alpha$  properly accounts for the fact that RTs for the male speaker are affected by a large number of preceding stimuli.

In this case, the `lmer` analysis is made more precise by using the DIANA-inspired detrending of the observed RT sequences to remove the local speed effect, before comparison takes place between the human RT sequences and the RT sequences simulated by DIANA.

## 6. Discussion and conclusion

This paper presents two simulations that show how data-oriented and process-oriented models can inform each other. In the first simulation, we show how a data-oriented `lmer` uncov-

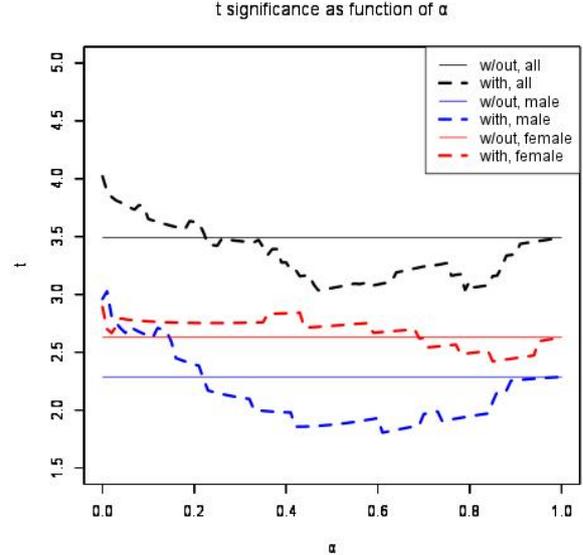


Figure 3: Values of the  $t$ -statistic for the male (spk A, in blue, lower curves), and female speaker (B, in red, middle curves), as well as for both (upper curves in black). Values below 1.95 relate to insignificance of the match/mismatch condition in the prediction of RT.

ered what can be considered as an important shortcoming of DIANA, namely that fact that DIANA cannot account for changes in parameters that occur within an experiment. At the same time it was shown that an `lmer` model can give accurate suggestions for how the adaptation should be done, and that the adaptations required for simulating exemplar effects can be implemented in DIANA.

The second simulation shows how detrending of human RT sequences by an AR filter that was designed in the context of the development of DIANA improves the resulting between-participant correlation and the significance of the crucial match-mismatch condition in an experiment. The  $t$ -statistic for that condition in an `lmer` model appears to depend strongly on how local speed is dealt with.

The dangers (e.g [27]) but also the benefits of data models (e.g. [18]) are well understood. Advantages and limitations of process models, however, are more difficult to capture. For the time being, it is impossible to simulate complex processes such as speech comprehension on all levels, starting at the neurological level (the neuron), up to semantic integration. As long as complete biological realism is not possible, process models must make numerous assumptions. But these assumptions may be informed by the knowledge derived from data models. We are in the process of updating DIANA, such that the model will be able to simulate on-line adaptation in a cognitively plausible manner.

## 7. Acknowledgments

This work was funded by an ERC starting grant (284108) and an NWO VICI grant awarded to Mirjam Ernestus.

## 8. References

- [1] W. Gerstner, H. Sprekeler, and G. Deco, "Theory and simulation in neuroscience," *Science*, vol. 338, no. 6103, pp. 60–65, 2012.
- [2] R. Whelan, "Effective analysis of reaction time data," *The Psychological Record*, vol. 58, no. 3, pp. 475–483, 2008.
- [3] A. Cutler, *Native Listening: Language Experience and the Recognition of Spoken Words*. MIT Press, 2012.
- [4] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org>
- [5] L. ten Bosch, L. Boves, and M. Ernestus, "Towards an end-to-end computational model of speech comprehension: Simulating a lexical decision task," in *Proceedings of Interspeech*, Lyon, France, 2013.
- [6] —, "Comparing reaction time sequences from human participants and computational models," in *Proceedings of Interspeech*, Singapore, 2014.
- [7] —, "DIANA: towards computational modeling reaction times in lexical decision in north american english," in *Proceedings of Interspeech*, Dresden, 2015.
- [8] D. Norris and J. McQueen, "Shortlist B: A Bayesian model of continuous speech recognition," *Psychological Review*, vol. 115, pp. 357–395, 2008.
- [9] O. Scharenborg, "Modeling the use of durational information in human spoken-word recognition," *Journal of the Acoustical Society of America*, vol. 127, pp. 3758–3770, 2010.
- [10] I. Hanique, E. Aalders, and M. Ernestus, "How robust are exemplar effects in word comprehension?" *The Mental Lexicon*, vol. 8, no. 3, pp. 269–294, 2013.
- [11] S. Goldinger, "A complementary-systems approach to abstract and episodic speech perception," *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany*, pp. 49–54, 2007.
- [12] C. McLennan, P. Luce, and J. Charles-Luce, "Representation of lexical form," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 29, pp. 539–553, 2003.
- [13] M. Ernestus and A. Cutler, "BALDEY: A database of auditory lexical decisions," *Quarterly Journal of Experimental Psychology*, vol. Advance online publication, 2015.
- [14] M. Usher, Z. Olami, and J. L. McClelland, "Hick's law in a stochastic race model with speed-accuracy trade-off," *Journal of Mathematical Psychology*, vol. 46, pp. 704–715, 2002.
- [15] C. McLennan and P. Luce, "Examining the time course of indexical specificity effects in spoken word recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 31, no. 2, pp. 306–321, 2005.
- [16] J. J. Lee and C. F. Chabris, "General cognitive ability and the psychological refractory period: Individual differences in the mind's bottleneck," *Psychological Science*, vol. 24, no. 7, pp. 1226–1233, 2013.
- [17] S. Dufour, A. Brunellière, and U. Frauenfelder, "Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials," *Cognitive Science*, vol. 34, pp. 489–507, 2013.
- [18] R. H. Baayen, J. van Rij, C. de Cat, and S. N. Wood, "Auto-correlated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models," in *Mixed Effects Regression Models in Linguistics*, D. Speelman, K. Heylen, and D. Geeraerts, Eds. Berlin: Springer, (to appear).
- [19] M. Ernestus and R. H. Baayen, "The comprehension of acoustically reduced morphologically complex words: the roles of deletion, duration, and frequency of occurrence," in *Proceedings of ICPhS, Saarbrücken*, 2013, pp. 773–776.
- [20] A. Kelly, A. Heathcote, R. Heath, and M. Longstaff, "Response time dynamics: evidence for linear and low-dimensional non-linear structure in human choice sequences," *The quarterly journal of Experimental Psychology Section A: Human Experimental Psychology*, vol. 54, no. 3, pp. 805–840, 2001.
- [21] D. E. Meyer and D. E. Kieras, "A computational theory of executive cognitive processes and multiple-task performance: part 1. basic mechanisms," *Psychological Review*, vol. 104, no. 1, pp. 3–65, 1997.
- [22] R. Ratcliff, "Group reaction time distributions and an analysis of distribution statistics," *Psychological Bulletin*, vol. 86, pp. 446–461, 1979.
- [23] R. Ratcliff and J. N. Rouder, "Modelling response times for two-choice decisions," *Psychological Science*, vol. 9, p. 347, 1998.
- [24] E. Wagenmakers, S. Farrell, and R. Ratcliff, "Estimation and interpretation of  $1/f^\alpha$  noise in human cognition," *Psychonomic Bulletin & Review*, vol. 11, pp. 579–615, 2004.
- [25] S. Brown and A. Heathcote, "The simplest complete model of choice response time: Linear Ballistic Accumulation," *Cognitive Psychology*, pp. 153–178, 2008.
- [26] T. L. Thornton and D. L. Gilden, "Provenance of correlations in psychological data," *Psychonomic Bulletin & Review*, vol. 12, pp. 409–441, 2005.
- [27] N. Silver, *The Signal and the Noise: The art and science of prediction*. London: Allan Lane, an imprint of Penguin Books, 2012.