**Our responses to reviewer comments (actionable parts are underlined), are shown in red.**

**Vasishth and Engelmann**

**Clearance reading**

I greatly enjoyed reading (most of) this book manuscript! I look forward to reading the final, published copy, using the results in my own research, and assigning portions of the book to my students.  
  
My detailed comments are below. As a summative comment, I would say that at times the writing loses the forest for the trees. This is probably a natural result of the fact that the individual chapters and sections originated in published papers, which are by their nature (1) internally motivated and (2) encapsulated. I think the overall framework introduced in Chapter 1 is strong, and that a revision that attends to connecting individual research findings to this framework will greatly increase the coherence of the book.  
  
  
**Chapter 1**  
  
Generally well done. The review of prior modeling efforts is relatively short for pragmatic reasons and focuses on prior work of greatest relevance to what comes later, making the treatment self-contained. This point is effectively made later, on p. 16 of section 1.6. It might be work commenting on the scope of the review earlier, to assure the reader.

**Response**:

We have added the following text in the opening paragraph of chapter 1:

“Because of the narrow focus of the work presented here, our discussion of alternative computational modeling approaches will be cursory. This is in no way intended to diminish the importance of these approaches; we feel that approaches such as connectionist modeling and non-linear dynamical systems based models add great theoretical value to the field and deserve a fuller treatment.”

The discussion the third gap the current book addresses – the data analysis gap that is one of Vasishth’s specialties – is particularly good. I can see assigning this segment alone to student working far outside of psycholinguistics.  
  
I wish that section 1.6.3 was a built bolder. I agree with the authors on the points made therein. It is essential to develop comprehension models of reading performance, and the book promises to do so in synthesizing and systematizing 15+ years of modeling work. That said, there are surely integrative lessons that were learned in the creation of this book – insights into where the different findings or model mechanisms don’t quite cohere, insights into which phenomena are particularly intractable for comprehensive (or ACT-R-based) models and therefore deserve further theoretical work, insights into how working within an architecture raises questions for the modeler when the architecture itself evolves over time, etc. I assume these insights are scattered throughout the rest of the book. Is the generation of such insights one of the consequences of the comprehensive modeling they have undertaken, and if so, is it is worth mentioning here?

**Response**:

This is a good point and in response to this concern we have added the following text in section 1.6.3. The additional text below attempts to address what we see as the reviewer’s main point: what lessons were learnt while carrying out the computational modelling?

“Because the modelling reported here was carried out over many years (most of the work was done between 2005 and 2018), many computational challenges had to be overcome. For example, the ACT-R architecture itself is continuously evolving independently of the sentence processing architecture we work with. These version changes in ACT-R necessitated a near-complete rewrite of the modelling constructs. As a consequence, the original Lewis and Vasishth model's underlying machinery also changed in subtle ways. This evolution of ACT-R will remain a challenge for future researchers. A further problem that we encountered was that lisp is not a widely used programming language any more; this makes the ACT-R model less accessible to researchers interested in using it. Fortunately, researchers have recently developed viable alternative implementations in python (Brasoveanu and Dotlacˇil, 2020) , which may be easier to maintain and develop further. The lesson to be learnt here is that developing a sustainable code base, and preventing code rot, is a major challenge in any large programming project like this one, and the user/reader needs to be aware of this limitation and to be patient when adapting or using legacy code. One insight to be gained here is that perhaps some compromises are necessary in order to make the theoretically machinery more accessible to the wider community. It is possible to investigate the core principles of the model without implementing a fully-fledged model; this can be done by using code written in the programming language R. We provide such an implementation, along with a Shiny app that allows the reader to compute simple effects without doing any coding at all.\footnote{https://engelmann.shinyapps.io/inter-act/}.

“Another issue we faced was that we studied different research problems piecemeal. For example, a model integrating eye-movement control and parsing is reported in Chapter 5, but this model has not been regression-tested with the core phenomena discussed in Chapter 4 or other chapters. Future generations working on this framework could (and should) develop a more systematic testing framework, so that empirical coverage is incremental in the sense that the model's performance on all previously modelled data-sets and phenomena is evaluated again when exploring an extension of the architecture.

“One further area where the present work fell short was that modelling should ideally always be comparative; a baseline model is necessary to evaluate a particular model's relative performance. In more recent work, reported in Chapters 7 and 8, we have attempted to shift the focus towards evaluating competing models' performance on the same data-set. In future work, this should be standard practice. For example, an alternative competing model of eye-movement control and parsing would be very useful in order to better understand the relative performance of the model presented in chapter 5.”

Chapter 2  
  
This chapter is very well written. The empirical phenomena of interest clearly introduced. In addition, the authors take a critical approach to the findings, holding them to the high empirical standards they introduced in Chapter 1, and they do so in a respectful manner. This is not often the case in the current replication crisis in psychology, where some researchers seem to take a certain glee in challenging the work of others. By contrast, the analysis here is objective and impersonal.  
  
I found the text in Figure 2.4 verging on too small. I could read it in my PDF copy by zooming in sufficiently. The reader of a paper copy will not have this luxury. If horizontal space is at too much of a premium, consider reorganizing as a vertical figure. (The same comment applies to Fig. 2.6.)

**Response**:

We have enlarged the figures.  
  
I lost my place in section 2.5.2, where the influence of different priors is considered. This felt “in the weeds” given what I thought was the goal of the chapter: to review the key empirical phenomena modeled later. I would recommend prefacing this section with a short paragraph explaining the relevance of the section to what’s to come. For example, it could be that you will be considering other priors below. If so, let the reader know. (Aside: You return to this topic briefly at the end of Chapter 3, on p. 68. But here, at least, you don’t make much use of the machinery developed in Chapter 2.)

**Response**:

We have now added a paragraph explaining why this section appears in a chapter summarizing the evidence available so far. The main point we explain is that data are always interpreted in the light of our prior beliefs. How can one formally (quantitatively) take one’s beliefs into account when interpreting data? In this subsection, we demonstrate how this can be done. The broader point here is to show that subjectivity and conflicting beliefs can influence what we believe the data tell us, and that’s fine---people have different priors beliefs which may be legitimate in their own right. This way of thinking is not standard in psycholinguistics, but it should be. As a first step towards achieving this goal, we demonstrate how to incorporate prior beliefs into the interpretation of data.

“In this chapter, we summarized the main evidence available from reading studies relating to interference effects in different dependency types. On the sur- face, one might think that empirical data are “objective” in the sense that they speak for themselves. However, in practice researchers always interpret data in the light of their prior beliefs, and sometimes these beliefs can be very strong. When these prior beliefs are strong, it makes intuitive sense that a single counterexample from one experimental result should not change our beliefs much. In the case of reflexives, in the course of informal discussions, researchers have expressed skepticism about the estimates of the facilitatory interference effect in English reflexives reported in Ja ̈ger et al. (2020). The argument here is that there is a lot of prior data that doesn’t match the Ja ̈ger et al. findings. Essentially, the objection is that the analysis of the Ja ̈ger et al. replication data, we do not use all available information from prior work. This kind of use of prior knowledge need not be invoked informally; one can simply take one’s prior beliefs into account in the data analysis. This is not normally done in psycholinguistics, but with the increasing availability of Bayesian tools for data analysis, it is easy to formally incorporate prior beliefs into account.

“In this section, we briefly demonstrate how prior beliefs can be taken into account in the specific case of the data from Ja ̈ger et al. Below, we illustrate how the expert can formally interpret available data in the light of either prior data or their own prior subjective beliefs. The reason we bring this point up here is that we feel that incorporating prior beliefs in the analysis is a very important tool for understanding “what the data tell us”. The data never “speak for themselves”, they always speak through the filter of our beliefs. We show below can this subjectivity be formally incorporated into the interpretation of data.”

Chapter 3  
  
Section 3.2.1 is awfully short. I could have used more exposition here. Without it, the Figure 3.3 caption was insufficient for me to understand what is being shown. Is it that for a range of Latency Factor and MAS values, the LV05 model produces the correct interference effect for the target match vs. mismatch conditions?

**Response**:

This section was included in error; it is actually from an earlier version of the book, and is no longer needed, and has been deleted. These predictions are discussed in what was Figure 3.5, later in the chapter.

Equation 3.6 repeats Equation 3.1 (with the trivial difference of “ln” vs. “log”). If this redundancy was unintended, then consider deleting Equation 3.6. However, if it was intentional, for example to help the reader remember this critical equation, then consider “instantiating” its general variables with those specific to the LV05 model, to aid the reader’s comprehension. (Same comment applies to Equations 3.7, 3.8, and 3.10.) On a related note, should Equation 3.9 be presented earlier, with the general ACR-R memory retrieval equations, or is it somehow specific to the LV05 model?  
  
p. 57 doesn’t have to define the fan effect again, except perhaps in a footnote.  
  
You had earlier mentioned a “race” process and I assumed you were just using the “horse race” metaphor for competition where there is only one winner. You had previously cited Logacev and Vasishth (2015) and Raab (1962), on pp. 29-31, but I had missed their importance. On p. 58, you explain precisely what you have in mind – a specific technical consequence of ACT-R probabilistic retrieval dynamics. This is very interesting! The text here provides more of the details but I wonder if further improvement is possible. This seems like an important idea to highlight, to get it into people’s theoretical toolkits. (And the race idea continue recurs below, e.g., p. 74.)

**Response**: to-do rewrite that chapter

pp. 60-61: In showing that the model is committed to differential predictions for the Target-Match and Target-Mismatch conditions, regardless of the values of several ACT-R parameters, you have the opportunity to revisit Roberts and Pashler (2000) critique here. The model, to its credit, cannot fit \*any\* data pattern with the right choice of parameter values. In other words, it is falsifiable. Kudos!  
  
It was not clear to me how section 3.3 fits into the argument that the book has been developing. Perhaps the first sentence of this section could be expanded to make clearer the motivation for this section. I also found the section difficult to follow. Some of the technical details were elided, so I had to take the authors’ word at several points where I would rather have been able to make inferences on my own. If there is an opportunity to add more exposition here, it would greatly benefit the reader.  
  
  
Chapter 4  
  
This chapter needs an introductory paragraph or two \*before\* section 4.1 that transition from the prior chapters and motivate the work described therein. What is the major empirical challenge that needs to be addressed? What are the candidate mechanisms on the table for handling this challenge, and why choose prominence and cue-confusion? Is the goal purely to increase the empirical content of the model (i.e., to better account for the data), or is it also to incorporate new mechanisms from other sentence models or from the broader memory literature or from other aspects of ACT-R to increase the theoretical content of the model?   
  
Section 4.1 articulates 1’, 2’, and 3’, and motivates them as more realistic of cognition. Were the prior 1, 2, and 3 simplifying assumptions that were made in 2005-6, that are now being generalized here? Or were 1, 2, and 3 consistent with the ACT-R of the time, and are 1’, 2’, and 3’ updates to reflect how the architecture has evolved over the past 15 years? (Aside: Section 4.1.3 makes clear that 3’ actually capitalizes on a continuous mechanism that ACT-R offered all along, but which was simplified to a discrete mechanism in the LV05 model. This is the sort of detail I wanted previewed in the beginning of the chapter.)   
  
In a few places, the journal article origins of this chapter remain. For example, there is a section called “General Discussion”, and the appendices make reference to “the present article”. Also, when R is mentioned, the R Core Team is cited. I am not sure this is necessary given that it was cited in an earlier chapter. And the General Discussion cites earlier papers rather than earlier chapters regarding the probably low power of the literature (p. 101). These seams should be removed.   
  
pp. 73-77 are \*tough\* for the reader, in part because some of the arguments are presented as lists of propositions but more because of the frequent references to Figures 4.11 and 4.12 in the appendices. The coordination burden is hard. That said, I recognize that this is a very complex argument the authors are making, and a matching level of complexity in the writing is to some degree unavoidable. Might it be possible to move Figures 4.11 and 4.12 into the main text of the chapter, so they appear nearer to where they are referenced? Or, might it be possible to let Figure 4.2 do all of the work here, and to simplify the text just enough to make sense of the rest of the chapter, and to push the detailed text that’s currently here into the appendices with Figures 4.11 and 4.12 for the curious reader? Right now, the interleaving is tough for the reader, and the complexity of the argument in this early part of the chapter, before the data and model results are shown, is also tough.  
  
Note that once the chapter gets past this section, I found it very clear. It nicely describes the patterns in the data and how these are accounted for by 1’, 2’, and 3’. The description of the extension of ACT-R to implement 1’ (prominence) and 3’ (multi-associative cues) is crystal clear.  
  
The point about pre-registration seems a bit out of place on p. 101 as this topic had not been brought up earlier. Consider deleting it here (or possibly moving it to Chapter 1 – though I would recommend simply deleting it).  
  
More generally, the wrapping up (“General Discussion) sections of this chapter go on a bit too long. Consider streamlining this material and focusing on the conclusions most relevant for the argument developed in the book.  
  
  
Chapter 5  
  
I would disagree with the first sentence of this chapter: “In language comprehension research, most of the evidence about the cognitive processes involved comes from the study of eye movements in reading.” That is not my impression from the literature, especially for “language comprehension” writ large. It is certainly not the case, for example, for the literature on text and discourse comprehension. Consider reducing the scope of this sentence (e.g., “sentence comprehension”).  
  
There are again a few places where the journal article origins of this chapter remain. For example, at the bottom of p. 119, the LV05 model is “introduced” again. By this point in the book, the reader is well acquainted with this model. This text should be changed to something like “building on the LV05 model described in previous chapters…”. A similar comment applies to the brief description on the model’s parsing architecture on p. 127. This can be condensed and pointers to the earlier description inserted. And again, when R is mentioned, the R Core Team is cited (p. 123). I am not sure this is necessary to repeat this in each chapter. Finally, the “Future prospects” section works less well as an ending to a chapter. It leaves the reader with a lot of (interesting!) questions. But it is preferable to end the chapter by connecting to the larger argument developed thus far in the book and by previewing how the argument will be continued in the (highly related) next chapter.  
  
To interpret Equation 5.1 (p. 121), it would be helpful to know the ranges on the k and epsilon parameters. I assume both must be non-negative (for the equation to make sense to me, conceptually) but perhaps I am wrong. (I suppose I am also making an assumption about how epsilon is measured which may be wrong. This should also be specified.)  
  
The text in Figure 5.1 (p. 123) is readable online with sufficient zooming, but I think it would be very hard to make out the legend text (and perhaps the markers) in a printed book. Consider increasing the size of this figure. (Same goes for Figure 5.2.)  
  
The legend for Table 5.2 defines \*some\* of the variables and abbreviations found there. However, I note there is less interpretation of the data than there is in the table legends and especially the figure captions of prior chapters. Consider including more of this here to maintain consistency of style across chapters.  
  
[I am through the first three sections of this chapter and I want to comment that I am greatly enjoying the clarity of the presentation. (And I am not an eye movements researcher.) And unlike in the prior chapter, the placement of some information in the appendix is not impeding the flow of my reading.  
  
I wonder, with such short subsections in section 5.3 (often one paragraph long!), whether this text should be organized more “flatly” than would be the case in a journal article.  
  
The text introducing Hale’s surprisal and setting it up as potentially complementary to the retrieval interference construct of LV05 is good, and so theoretically interesting (p. 119; 127-128)! As currently written, it is a bit backgrounded to the chapter’s goal of bolting on Emma to LV05. I wish it could be foregrounded more. This is a nice example of the benefits of working within an architecture (i.e., model integration). It is also an even-handed evaluation of competing (or complementary) theoretical proposals. I find it even more interesting than the engineering goal that seems to drive the chapter! In this vein, I also find the brief discussions of the seams between the two models, and the possibilities for the future research they generate, to be theoretically interesting. They are again examples of the wins that come from working within an architecture and for constructing comprehensive accounts of the word and sentence comprehension. I would have appreciated even more discussion of these matters. (Looking ahead, Chapter 6 seems to take on the matters of architecture, model integration, and model sufficiency, so perhaps it is OK to defer on these matters here in Chapter 5.)  
  
It might be useful to include a timing figure to illustrate the sequencing of actions during the time-out and exit-time out phases of a regression, and how these differ from the sequencing during normal reading.  
  
  
Chapter 6  
  
Figure 6.1 (p. 148) is difficult to understand on its own, with its heavy use of abbreviations. Could “FFD”, “FPRT”, and “RPD” be replaced with the phrases they abbreviate in the figure itself, and could the other abbreviations be defined in the caption?  
  
p. 152: The way Kemper et al. (2004) is described here, it seems like a replication of MacDonald et al. (1992, Cognitive Psychology) with respect to the working memory finding. If I have this correct, consider citing the earlier paper.  
  
It may be possible to re-organize section 6.3 (specifically, 6.3.1) to have psycholinguistics questions join the modeling question in motivating this research. Specifically, the two full paragraphs on p. 155, which come at the end of the literature review, nicely lay out the psycholinguistic questions around under-specification and good-enough parsing, especially for low WM readers. Consider introducing these questions in briefer form at the beginning of the section to motivate the literature review that follows.  
  
As an aside, I found the modeling in section 6.3 really insightful! I was left with questions about how the modeling of the impact of WM differences on the processing of ambiguity connected to the capacity-constrained architecture of Just and Carpenter (1992) and to the beam search size of Jurafsky (1996). Consider whether it’s worth commenting on these potential connections.  
  
  
Chapter 7  
  
It would be useful if the introduction to this chapter tied its contents to the overall structure of the book in a bit more detail than the paragraph on p. 164.  
  
Consider connecting the discussion of the SRC vs. ORC difference in section 7.2.2 (pp. 170-171) to discussions of this difference in the prior chapter. This would also present the opportunity to discuss in greater detail the relationship between the LV05 model (+ Emma extension) model presented earlier (and called M0 here?) and the mixture/STAN-based accounts presented in chapter 7. I feel like I am missing a broader, more integrated understanding of the various approaches to the  SRC vs. ORC difference in the different chapters of the book. It is probably there implicitly; as the reader, I just need a little help seeing it.  
  
The chapter is presented as contrasting different approaches to interference during sentence comprehension – those offered by McElree and by LV05 – in the introductory paragraph on p. 164. But the reader notices quite quickly in section 7.2 that a more neutral statistical model is being offered, and it is couched in a different computational formalism (STAN) than the models thus far (ACT-R, R). Then, at the bottom of p. 175, we get this statement of the goal of section 7.3: “Our goal here is to implement all the three proposals as statistical models and then compare their relative fit to the data in order to adjudicate between them.” I actually believe that this is closer to the goal of the chapter – the formulation of some of the (earlier) ideas in statistical terms and their direct evaluation. This is certainly the spirit of the modeling thus far in the chapter, as shown in the model specifications (Equations 7.1 – 7.5), which are quite different than what has come before. If I am correct, it may be worth expanding the introduction to this chapter to offer this as another motivation for the research reported in this chapter.  
  
The chapter could use a summary section 7.4 that connects back to the top-level research questions (interference accounts, statistical modeling) that motivated it.  
  
  
Chapter 8  
  
I really like the way this chapter begins and motivates the research it reports. The review of the literature in section 8.1 is both comprehensive and highly readable.  
  
The text at the beginning of section 8.2 seems a bit redundant with what has come before. First, it summarizes the representation vs. processing distinction that section 8.1 covered in depth (and, as an aside, refers to “this paper” on p. 191). Then it summarizes the LV05 model that the reader has already encountered in previous chapters. Consider cutting back some of this material.  
  
[I ran out of time and could not review the rest of this chapter.]  
  
  
Chapter 9  
  
This chapter is highly schematic so I did not review it in great depth. As the authors develop it further, they might find it useful to organize it around the research questions and goals introduced in sections 1.5 and 1.7. What has the reader learned about these from the intervening chapters, and what issues remain to be addressed? Note that such an organization would not naturally accommodate section 9.4, which seems independent of the rest of the book. As such, it might belong in a section considering computational modeling of sentence comprehension more broadly.  
  
  
Miscellaneous Notes  
  
The authors cite Just et al. (1999) with respect to 4CAPS and the sentence comprehension model developed within in. More comprehensive descriptions are found in Just and Varma (2007, Cognitive, Affective, and Behavioral Neuroscience) and Varma (2016, Oxford Handbook chapter).  
(1) Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of cognition. Cognitive, Affective, and Behavioral Neuroscience, 7, 153-191.  
(2) Varma, S. (2016). The CAPS family of cognitive architectures. In S. E. F. Chipman (Ed.), The Oxford Handbook of Cognitive Science (pp. 49-68). Oxford University Press.  
Apologies for suggesting my own papers here!  
  
The citation for Logacev and Vasishth (2015) is missing information.  
  
Figure 4.1 is referenced in chapter 3, on p. 57, but does not appear until chapter 5, on p. 71.  
  
p. 70: “constitutes” ◊ “constitute”  
  
p. 71: “certain representation” ◊ “a certain representation” or “certain representations”