Draft

Computational models of retrieval processes in sentence processing

Shravan Vasishth
Department of Linguistics, University of Potsdam, Potsdam, Germany

Felix Engelmann
Manchester University, UK

Bruno Nicenboim
Department of Linguistics, University of Potsdam, Potsdam, Germany

Frank Burchert
Department of Linguistics, University of Potsdam, Potsdam, Germany

June 13, 2019

Abstract

Sentence comprehension requires that the comprehender work out who did what to whom. This process has been characterized as retrieval from memory. We review the quantitative predictions and empirical coverage of computational implementations of retrieval models. Four directions are identified for future work: competing computational models should be developed so that their relative performance can be tested against common benchmark data-sets; large-sample, cross-linguistic data-sets need to be created to allow for quantitative model comparison; computational modeling of sources of variability in both unimpaired and impaired sentence comprehension needs to be carried out; and a shift in emphasis is needed towards accounting for systematic variability at the individual-participant level.

Keywords: cue-based retrieval; similarity-based interference; comprehension impairments in aphasia; individual differences

Memory retrieval in sentence processing

Models of language processing in adult native speakers have from the very beginning assumed that working memory constraints play a role in determining processing difficulty. Some examples: (a) In sentence processing theories, working memory has been to motivate

Correspondence: vasishth@uni-potsdam.de.

through heuristic principles like Late Closure and Minimal Attachment [1], which favor the creation of simpler structures over more complex ones for reasons of storage efficiency. (b) Working memory constraints and individual differences have been investigated systematically (and not without controversy [2]) in some of the early influential studies involving reading [3]; these topics remain an active area of research today [4, 5, 6, 7]. (c) The arguments and discussions regarding the parsing algorithms used by the human sentence processing mechanism were informed by space complexity considerations [8]. (d) Several researchers [9, 10] have proposed that minimizing distance between co-dependents (such as a subject and a verb) in a sentence facilitates comprehension and leads to easier processing (for evidence in favor of distance minimization, see [11, 12, 13, 14, 15, 16]). The tendency to minimize distance between co-dependents can be seen as arising from a need to reduce *similarity-based interference* from other items that are being held in memory [17].[1] Similarity-based interference has been used in cognitive psychology to describe the increased difficulty we experience when trying to recall a particular target item from memory, when other items that share some similarity with the target item are also being held in memory.

[24, 25, 26] were the first to systematically draw out a connection between similarity-based interference research in cognitive psychology and retrieval processes in sentence processing. Extending these ideas, [27] developed a computational model that furnishes quantitative predictions for interference effects reported in the literature. The empirical evidence for interference that emerged in subsequent years seems to show some evidence supporting the idea that constraints on working memory are responsible for certain crucial aspects of sentence comprehension difficulty [5, 17, 28, 29, 30, 31]. However, a recent meta-analysis of approximately 100 reading experiments [32] suggests that the overall evidence in favor of the role of interference in sentence processing is mixed at best.

Here, we review the empirical and theoretical insights that have recently emerged from computationally implemented models of retrieval processes; these are the activation model [27, 33] and the direct-access model [34, 35]. We focus mainly on computational (rather than verbally-stated) models of retrieval because they furnish predictions whose fit to data can be quantified. We also discuss some of the important open questions that need to be tackled in future work.

A computational model of cue-based retrieval

Although several models of retrieval exist, we start by discussing a sentence processing model [27]—hereafter, the activation model—that was implemented within the ACT-R cognitive architecture [36]. ACT-R (http://act-r.psy.cmu.edu/) is intended to serve as a general framework for modeling cognitive processes in different domains; it is used in many different areas, such as human factors research and cognitive psychology, to model behavioral data from a variety of experimental paradigms.

An accessible overview of the essential principles behind the activation model is presented in [33]. This model has been used to investigate a broad range of phenomena:

[1] There are several systematic counterexamples to distance minimization [13, 18, 19, 20, 21, 22] but these can be explained in terms of expectation effects driven by assuming that readers use a predictive, probabilistic, left-corner parsing mechanism and that their degree of surprise upon encountering a word is a function of probabilistic expectations. Expectations about upcoming input based on probabilistic knowledge about the language appears to be a largely independent factor influencing comprehension difficulty [23].

similarity-based interference effects [37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]; the relative roles of predictive processing vs. interference effects [49]; impairments in individuals with aphasia [50, 51, 52]; the interaction between oculomotor control and sentence comprehension [53, 54]; and the effect of working memory capacity differences on underspecification (good-enough processing [55]) in sentence comprehension [56]. Because this computational model has been extensively investigated and has served as a baseline for several new emerging computational models [57, 58, 59, 60], we begin by summarizing its key predictions, and the empirical evidence both for and against the model.

Some key assumptions of ACT-R and the activation model

ACT-R incorporates several assumptions about working memory processes that are potentially relevant for sentence processing. A central assumption is that any information processing task consists of manipulating, combining, changing, or creating information units in memory that are called *chunks*; these transformations on chunks are carried out via a *production system*, which is a series of if-then condition-action statements. Every chunk has associated with it a unit-less real-value number—its activation—which determines how easy it is to access from memory. The higher the activation, the easier it is to access the chunk. Every chunk has some baseline activation value, and over time the activation of a chunk experiences exponential decay, following the power law of forgetting [61]. Apart from decay, activation is also affected by a random noise component. A key component of ACT-R is the *spreading activation* equation, which induces retrieval interference. Below, we explain spreading activation informally, using an example.

The activation model predicts two main classes of effect that fall out of the ACT-R architectue: inhibitory interference and facilitatory interference. These are discussed next, with quantitative predictions shown in Figure 1.

Inhibitory interference effects. In the course of carrying out an information processing task, if a previously processed chunk needs to be accessed, a parallel search is carried out in memory using a set of so-called *retrieval cues*. Retrieval cues are feature specifications; for example, the cue {singular} seeks out a chunk in memory that has singular number marking. The search mechanism is based on the well-established assumption in cognitive psychology (among others, [62, 63, 36, 64, 34, 35, 65, 66]) that memory access uses a cue-based content-addressable mechanism.

An example, adapted from [40], illustrates the retrieval process. Consider these sentences:

(1) The bodybuilder$^{+singular}_{+subject}$ who met the trainer$^{+singular}_{-subject}$ was$\{^{singular}_{subject}\}$ ...

(2) The bodybuilder$^{+singular}_{+subject}$ who met the trainers$^{-singular}_{-subject}$ was$\{^{singular}_{subject}\}$ ...

In (1), at the auxiliary verb *was*, an attempt is made to retrieve the singular subject noun phrase (*the bodybuilder*), but an intervening noun phrase (*the trainer*) also has the feature {+singular}. Whenever two or more chunks have a feature that matches a retrieval cue, the activation of the chunks is attenuated compared to the case where a unique chunk has a feature that matches the cue (the latter situation is exemplified in 2; the retrieval cue {singular} matches only the first noun). As a result of the attenuated activation in (1), and regardless of which chunk is retrieved at the auxiliary, a slowdown is predicted at the

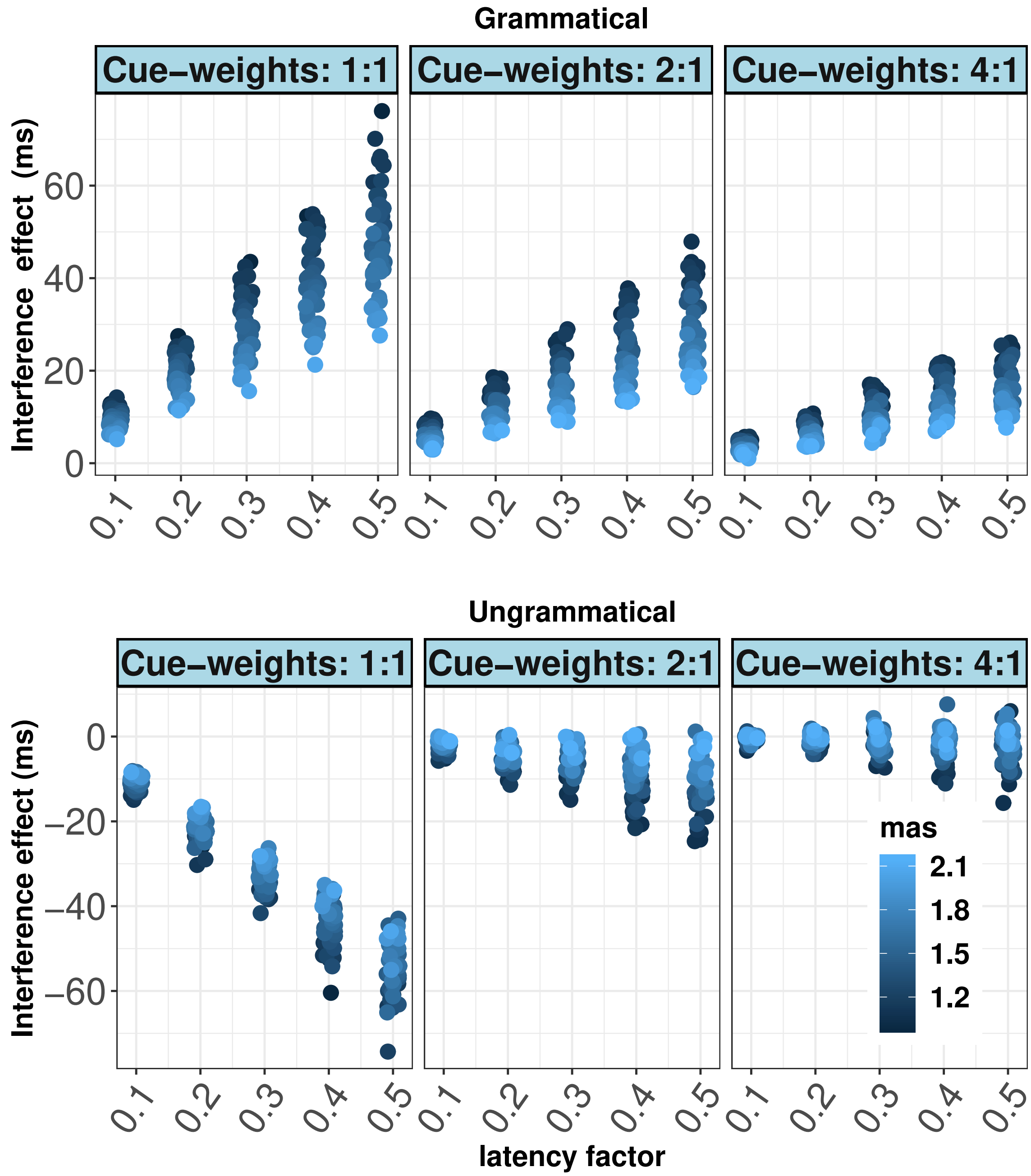

Figure 1. Shown are inhibitory and facilitatory interference effects in the activation model [27] (also see Box 1). The quantitative predictions are shown for different values of three parameters in the model: latency factor, which represents processing speed; maximum associative strength (mas), which controls the strength between retrieval cues and features of chunks in memory; and cue-weighting, which controls whether a particular cue has a higher weight than another cue (expressed as a ratio). Note that the interference effects (inhibitory and facilitatory) shrink toward 0 as the cue-weight increases; this is discussed further in the section Modeling Individual Differences.

auxiliary verb, relative to a baseline condition (2) where the intervening noun (*trainers*) does not have the feature {+singular}.

Although there is evidence for such interference effects from reading studies [28, 29, 30, 31], several studies report an absence of significant inhibitory interference effects [40, 48, 67, 68, 69, 70, 71] and some [7, 41] even report a facilitation where a slowdown is expected. The null results could certainly mean that retrieval does not play a major role in grammatical sentences; but there are other explanations [32], low statistical power in these studies being a plausible candidate [47]. The facilitation observed in [7, 41] could indicate that the predictions of retrieval theory are wrong; alternatively, the speedups observed in reading could be due to low working memory capacity participants engaging in good-enough processing [55] and and engaging in partial structure building during parsing.

This predicted slowdown in retrieval time in (1) vs. (2) is sometimes referred to as *inhibitory interference* [32, 72]. In ACT-R, inhibitory interference is implemented via the spreading activation equation mentioned earlier. As the name suggests, spreading activation attenuates the activation associated with each chunk whose features match the retrieval cues. The relative weights of individual cues can modulate the magnitude of the interference effect; this becomes important in the discussion below. Beyond ACT-R, other theories of retrieval in sentence processing [28, 35] also predict inhibitory interference effects (we return to a computational implementation of one of these theories below).

Facilitatory interference effects. The cue-based retrieval mechanism in ACT-R assumes that if a subset of the retrieval cues matches the features on a chunk in memory, the probability of the chunk being retrieved increases compared to when no retrieval cue matches the chunk's features. Now, if the features on one chunk match only some subset of the retrieval cues and the features on a second chunk match a *different* subset of retrieval cues, the model predicts a speedup in retrieval time (compared to a baseline condition where only one chunk matches a subset of the retrieval cues). The explanation for the predicted speedup from the activation model is discussed below (there are other plausible explanations, such as local coherence [73, 74]).

As an example, consider the sentences below, which have been investigated extensively in reading studies [40, 67, 68, 69, 71]:

(3) *The bodybuilder$^{-plural}_{+subject}$ who met the trainers$^{+plural}_{-subject}$ were$\{^{plural}_{subject}\}$ ...

(4) *The bodybuilder$^{-plural}_{+subject}$ who met the trainer$^{-plural}_{-subject}$ were$\{^{plural}_{subject}\}$ ...

Unlike (1, 2), both the sentences above are ungrammatical (marked with an asterisk following linguistic convention) because the auxiliary verb seeks out a plural-marked subject noun phrase, but the subject (*bodybuilder*) is singular marked, so it matches only one retrieval cue (the subject cue). As in example (3), if another plural marked noun (here, the noun *trainers*, which cannot be a subject because it is the object of a relative clause) matches the other retrieval cue (plural), then faster retrieval time is observed at the auxiliary verb, compared to a baseline condition (4) where the distractor (*trainer*) has singular marking and therefore doesn't match either of the two retrieval cues.

Withing the activation model, the underlying cause for the observed speedup is a so-called *race process* [75]: one process is an attempt to access one chunk, and the other process is an attempt to access the other chunk; both these processes unfold in parallel.

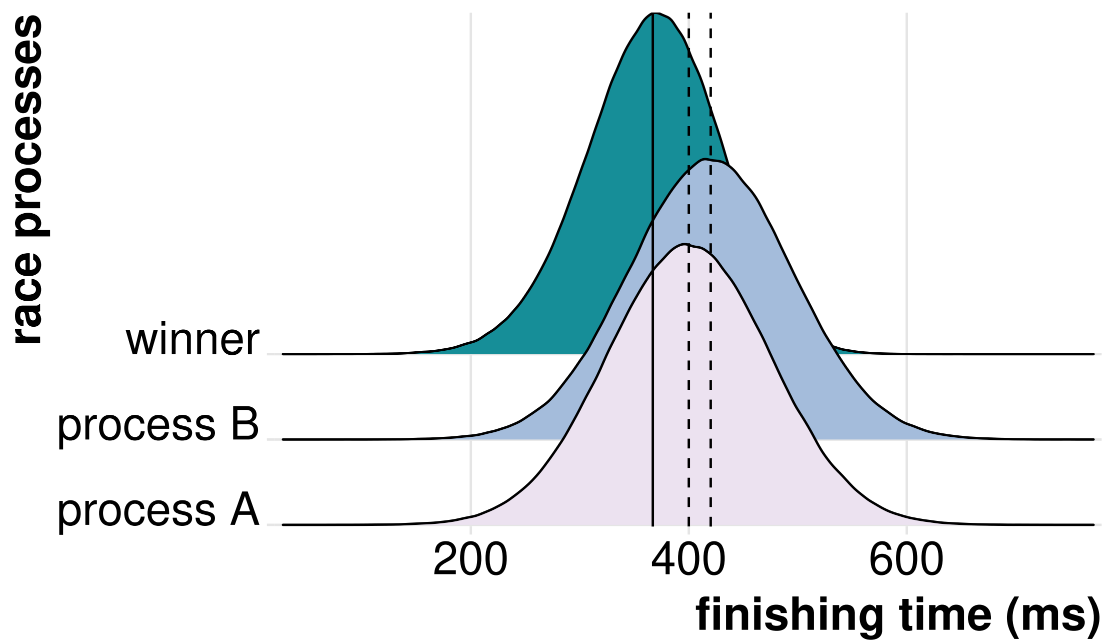

Figure 2. An illustration of a race process; here, we show the resulting reading time distribution that arises from a race between two processes (e.g., one process is an attempt to retrieve one chunk based on a set of retrieval cues, and the other process, triggered simultaneously, is an attempt to retrieve a different chunk). The mean of the winning finishing time distribution (the solid vertical line) is smaller than the means of the two processes (broken vertical lines) that are engaged in the race: the winning distribution represents the observed statistical facilitation.

Whichever process finishes first leads to the respective chunk being retrieved. The race process has the effect that on average, a faster reading time is observed compared to a baseline condition where no race occurs. In the above example, (3) is read faster than (4). The reading time distribution that results from a race process is illustrated in Figure 2. To our knowledge, of all the retrieval theories that have been proposed for sentence processing, only the activation model predicts facilitatory interference effects.

The activation model is not the only way to characterize retrieval processes in sentence comprehension. An alternative model, the direct-access model, is discussed next.

A computational implementation of the direct-access model

An influential alternative to the activation model is the direct-access model [34, 35, 65, 66]. In the literature on retrieval processes, the activation model and the direct-access model have been considered to be notational variants of each other [33]. Indeed, both models assume that retrieval is driven by a content-addressable memory process. However, the direct-access model does not predict facilitatory interference; this is because the underlying retrieval process assumed in the direct-access model is quite different from the activation model, as we explain below.

[37] noticed that the activation model can be seen as a lognormal race of accumulators

of evidence: given two or more alternative chunks that could be retrieved, the activation of each chunk determines which one ends up being retrieved in a particular trial. The activation acts as a rate of accumulation of evidence for each of the alternative structures [76, 77, 78]. The retrieval time in the activation model is determined by this lognormal race process.

By contrast, the direct-access model of McElree assumes that when a cue-based retrieval is carried out, only the probability of successful retrieval is affected, not the retrieval time per se. Any slowdowns observed in reading time are an artefact of occasional failures to retrieve the correct target, followed by an attempt to re-retrieve that element (a reanalysis step). [37] showed that successful retrieval in the direct-access model can be seen as a two-component mixture process [79], with some proportion of trials representing a fast, successful retrieval in the first-attempt, and some proportion representing a slower retrieval that results from an unsuccessful retrieval in the first attempt and a subsequent reanalysis step (for related work, see [80]).

[37] used the probabilistic programming language Stan [81] to implement the direct-access model as a Bayesian hierarchical finite mixture process, and the activation model as a hierarchical lognormal race of accumulators. They showed that at least for one data-set [47] with a relatively large-sample (182 participants) showing inhibitory interference effects, the direct-access model exhibited a better predictive performance (as measured by k-fold cross validation [82, 83]) compared to the lognormal race model. The reason that the direct-access model outperformed the activation model is that the activation model predicts that in inhibitory interference experimental designs, retrievals of the incorrect chunk should be slower than the retrievals of the correct chunk. However, in the data-set used for modeling [47], incorrect retrievals had faster reading time than correct retrievals.

Implementing these two competing models of retrieval processes made it possible to understand their relative predictive accuracy. This demonstrates one important advantage of implementing theories computationally: their predictive performance can be quantitatively compared.

One limitation of the [37] evaluation was that the model comparison was carried out on a single (albeit large-sample) data-set. An important research direction for the future is a broader evaluation of the relative predictive performance of the activation model and the direct-access model, as well as a comparison with other competing models that have emerged recently [57, 58, 59, 84]. Quantitative model comparison of multiple competing models across a wider spectrum of benchmark data is likely to yield new insights into the underlying retrieval processes in sentence comprehension. The first steps in expanding the model comparison to other data-sets have already been taken [52, 80]. A further important research area is the evaluation of the relative performance of competing retrieval models in explaining impairments in sentence comprehension [52]. Below, we discuss the recent work relating to modeling impairment.

Modeling impairments in sentence comprehension

One test of any theory of sentence processing is whether it can account for processing deficits in populations with impairments: if particular constructs within a sentence processing theory can be shown to be related to specific deficits in comprehension, this increases the plausibility of those theoretical constructs. Using a model of unimpaired processing to account for deficits in sentence processing can also yield insights into what the underlying causes of the observed deficits might be. Sentence comprehension difficulties in aphasia are an important example.

Aphasia is an acquired language disorder caused by a brain damage that can affect speech production and comprehension to varying degrees.[2] Several theories have been proposed to explain the nature and sources of comprehension deficits in individuals with aphasia [85]. The activation model has been used to characterize different sources of comprehension difficulty by varying specific parameters that have theoretical interpretations in the context of aphasia [50, 51]. This work shows that the behavior of each individual with aphasia can be characterized as being affected to varying degrees by three types of deficits that have been independently proposed in the literature on aphasia.

Some important future directions in modeling comprehension impairments in aphasia are the following: (a) Expand the cross-linguistic empirical base of the investigation. [51] relied on the available data from English, but large-sample data-sets from other languages are needed: the diversity of grammatical constraints, word orders, and morphosyntactic cues available in different languages are known to lead to different processing strategies in unimpaired populations [14, 15, 21, 22], and such factors may play a role in aphasia too [86, 87]. (b) Evaluate the relative predictive performance different (retrieval) theories using data from individuals with aphasia and controls. For example, recent work [52] compares the predictions of the activation model and the direct-access model to the English relative clause data from IWA and controls, and shows that the activation model furnishes better predictions than the direct-access model. This kind of model comparison is needed for other recently developed models [57, 58, 59, 60] that are intended to serve as alternatives to the activation and direct-access models.

Concluding remarks

Apart for the need to model impairments in sentence processing, we see two other important challenges for future work on retrieval processes. The first is the lack of appropriately powered data; without precise empirical estimates, it is very difficult to evaluate the predictions of quantitative models. For example, in [88] the authors attempted to compare the activation model's predictions against the 100 or so data-sets in [32]. However, the uncertainty of the estimates from the published data-sets is relatively high, making it difficult to draw firm conclusions [89]. The general lack of publicly available data [90] and near total absence of model source code are also barriers to model evaluation.

The second open issue is that, given the fact that computationally implemented models of retrieval processes exist, a great deal of insight is potentially being missed by focusing only on modeling average differences; individual-level effects are equally informative. We discuss these points next.

The absence of appropriately powered benchmark data

In their classic article [89] on testing model predictions, Roberts and Pashler point out an important barrier to evaluating predictions of quantitative models: the data need to be sufficiently precisely estimated. In other words, the statistical power needs to be sufficiently high (a common recommendation is 80% or higher [91]). As [32] show, it is likely that many

[2]See https://www.asha.org/public/speech/disorders/aphasia/.

of the published studies on retrieval processes in sentence comprehension have relatively low power. Psychologists [91, 92] have long pointed out the importance of high statistical power for making discovery claims, but these recommendations have not generally been adopted in psychology and psycholinguistics.[3]

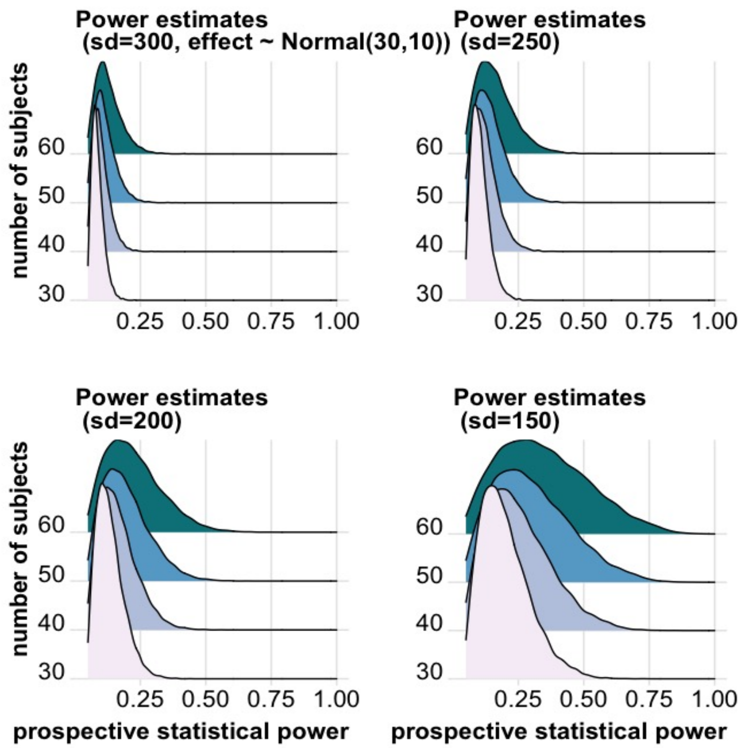

Figure 3. Power estimates for three subject sample sizes (30, 40, 50, 60), assuming a standard deviation (of the residual) of 150, 200, 250, 300 (a typical range in reading studies), and an effect size represented as a Normal distribution with mean 30 ms and standard deviation 10 (i.e., the effect ranges from approximately 10-50 ms with probability 0.95). For a justification of these estimates for the sample size, standard deviation, and effect sizes, see [32].

Power estimates (based on the meta-analysis in [32]) for reading studies on retrieval are shown in Figure 3; for typical effect sizes (10-50 ms), and commonly seen standard deviations (150-300 ms) in reading studies (self-paced reading and eyetracking), and routinely used participant sample sizes (30-60),[4] estimates of power are generally well below 80%.

There are many adverse consequences of publishing experiments with relatively low power (most experiments investigating retrieval fall within the 6-50% range). [96] demonstrate these consequences by carrying out seven replication attempts of a published paper [23].

[3] In response to the replication crisis that (partly) resulted from underpowered studies [93], several remedies have been suggested, such as reducing Type I error to 0.005 [94], or abolishing statistical significance testing entirely [95]. But in psycholinguistics, and indeed in any experimentally oriented research program, there is no substitute for an adequately powered study, and direct replications, if one wants robust findings [96].

[4] In repeated measures designs, experimental items also contribute a source of variance; power can be calculated for such designs using simulation. Simulation-based power calculations, taking into account item variability, yield power estimates similar to the ones presented here [48, 96].

Briefly: (a) under repeated sampling, most findings from underpowered studies will be null results from which nothing can be concluded; (b) any statistically significiant experimental results will tend to not replicate (i.e., statistically significant effects will tend to not come out significant in replications); (c) the uncertainty of the estimates will be very high, making a broad range of quantitative claims consistent with the data; and (d) any statistically significant effects in severely underpowered studies (under 20% power) will be gross misestimates of the true effect, due to Type M and/or S error [97].[5]

The main problem raised by underpowered studies for evaluating quantitative model predictions is that a large uncertainty associated with empirical effect estimates will lead to weak conclusions about model fit [89]: e.g., if the mean of an empirical estimate of an effect matches a model's quantitative prediction, but the 95% confidence interval of the mean spans a large range of possible values, then we will have high uncertainty about whether the model prediction matches the observed data.

In sum, what is urgently needed is a recognition that prospective power analyses be carried out before conducting a study, and that we ensure that our conclusions are based on appropriately powered experiments. In many cases, due to logistical limitations, it may not be possible to carry out high-powered studies; in those situations, Bayesian data analysis methods become necessary [98, 99].

Modeling individual differences

"*The average is an abstraction. The reality is variation.*"

Spiegelhalter and Blastland [100, p., 289]

It is standard in psycholinguistics to focus on the average effect. We usually ask questions like: is the effect of interest present or absent? If our conclusion from the data is "effect present", the claim is that participants on average show the effect. If our conclusion is "no effect present", then the claim is that participants on average do not show the effect [101]. However, even ignoring the statistical problems with this reasoning [102, 103], theoretically important individual-level differences can lie behind both significant and non-significant effects.

It is not common practice to evaluate model predictions against individual-level data. This is surprising because many researchers—a recent example is [104]—have pointed out the importance of individual differences research in language processing, and specifically in sentence comprehension research. In another example, [105] write: "individual differences research can aid in advancing the architectural specification of the systems responsible for language, thus fostering more mechanistic explanations of the processes underlying language comprehension." At least in sentence processing, especially reading research, these observations have had relatively little impact.

Recent developments in Bayesian probabilistic programming software have made it much easier to flexibly specify hierarchical models that can incorporate different assumptions about individual-level variability [106, 107]. These advances in statistical modeling have opened up new possibilities in the evaluation of model fits at the individual level.

[5]Type M(agnitude) error refers to the fact that, under repeated sampling, statistically significant effects in low-power studies will be larger than the true value of the effect; and Type S(ign) error refers to significant effects sometimes having the wrong sign.

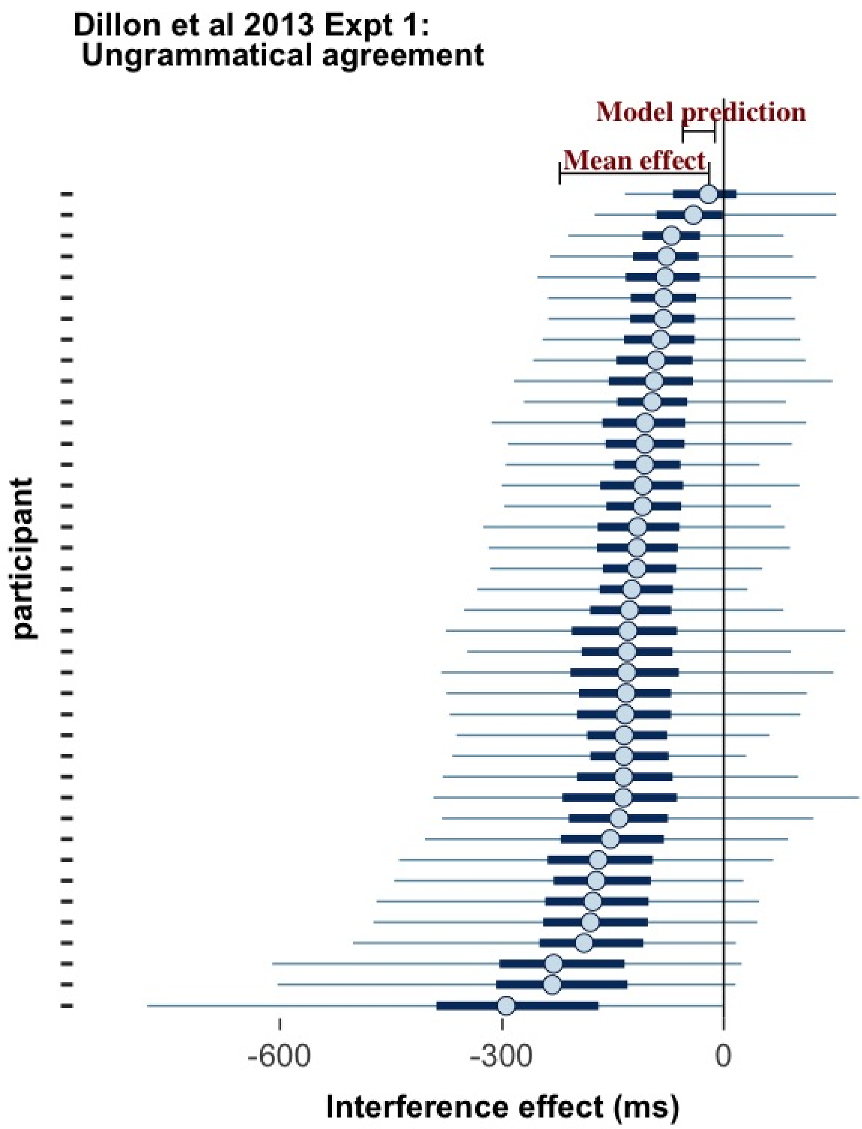

Figure 4. Shown are the individual-level facilitatory interference effects (circles show means, the thick lines 80% credible intervals, and the thin lines 95% credible intervals) in subject-verb agreement configurations; the data are from [40]. The participant-level estimates of the interference effect were computed from a hierarchical Bayesian model [112]. We see the predicted facilitatory interference effect even at the individual participant level, but different participants show varying magnitudes. The activation model can account for this variation in effect magnitude as a function of cue-weighting.

Example illustrating the theoretical importance of individual differences. Recall the subject-verb dependencies such as those shown earlier in examples (3) and (4). This is the configuration which shows facilitatory interference effects. Figure 4 shows the activation model's predicted effect, and estimates of the mean interference effect (means and 95% confidence intervals). Also shown are the participant-level interference effects (in milliseconds).

Some interesting observations emerge. First, we see a consistent facilitatory interference effect across participants; however, the magnitudes of the effects are larger in some participants and approach zero in others. The pattern observed here suggests that, within the activation model's framework, different participants could have different cue-weightings. Some participants may weight the subject cue higher than the number; these participants can efficiently target the subject without experiencing interference from the distractor noun. Others show large facilitatory patterns, which are consistent with equal cue-weighting. The activation model's cue-weighting parameter can explain this variation. Variation in cue-weighting between individuals could in principle be independently motivated by quantifying, for example, the linguistic skill (acquired through varying degrees of exposure) of the comprehender [108].

Notice that the activation model's predictions (Figure 1) for the pattern of individual differences are falsifiable; the model predicts varying degrees of mean inhibitory and facilitatory interference (modulated by cue weight) that is bounded at 0 ms. If data from multiple studies show that individual participants are clearly showing inhibitory interference when facilitation is expected (or vice versa), that would falsify the activation model; for discussion on evaluating patterns of individual differences, see [106].

Such a differentiated, individual-differences account cannot be developed if we focus only on average effects. As the quote from [100] above implies, the average is an abstraction that leads to generalizations masking theoretically important variation at the individual level. In future work, much theoretical insight can be gained by investigating individual-level effects in the context of quantitative model predictions.

In closing, although it is clear that working memory is not the only factor that determines processing difficulty in sentence comprehension [84, 109, 104], there is evidence supporting a role for retrieval in memory. Future work should aim to develop alternative computational models that can further help us understand the processes underlying retrieval events in parsing. This review covered the main recent empirical and theoretical developments in this research area, and discussed several important future directions. These are summarized in the box Outstanding Questions.

Acknowledgements

The research reported here was partly funded by the Volkswagen Foundation through grant 89 953; and the Deutsche Forschungsgemeinschaft (German Science Foundation), Collaborative Research Center - SFB 1287, project number 317633480 (*Limits of Variability in Language*) through project B2 (PIs: Shravan Vasishth, Frank Burchert and Nicole Stadie) and B3 (PIs: Ralf Engbert and Shravan Vasishth). We are grateful to Garrett Smith, Dorothea Pregla, and Sol Lago for comments on drafts of this paper. Thanks go to Brian Dillon for generously releasing his published data.

References

[1] Lyn Frazier. “On Comprehending Sentences: Syntactic Parsing Strategies.” PhD thesis. Amherst, MA: University of Massachusetts, 1979.

[2] David Caplan and Gloria S. Waters. “Verbal working memory and sentence comprehension.” In: *Behavioral and Brain Sciences* 22 (1999), pp. 77–126.

[3] Jonathan King and Marcel A. Just. “Individual differences in syntactic processing: The role of working memory.” In: *Journal of memory and language* 30.5 (1991), pp. 580–602.

[4] Ian Cunnings and Claudia Felser. “The role of working memory in the processing of reflexives.” In: *Language and Cognitive Processes* 28.1-2 (2013), pp. 188–219.

[5] Julie A. Van Dyke, Clinton L. Johns, and Anuenue Kukona. “Low working memory capacity is only spuriously related to poor reading comprehension.” In: *Cognition* 131.3 (2014), pp. 373–403.

[6] Titus von der Malsburg and Shravan Vasishth. “Scanpaths reveal syntactic underspecification and reanalysis strategies.” In: *Language and Cognitive Processes* 28.10 (2013), pp. 1545–1578.

[7] Bruno Nicenboim et al. “Working memory differences in long distance dependency resolution.” In: *Frontiers in Psychology* (2015).

[8] Philip Resnik. “Left–corner parsing and psychological plausibility.” In: *Proceedings of COLING*. address not known, 1992, pp. 191–197.

[9] Marcel A. Just and Patricia A. Carpenter. “A capacity theory of comprehension: Individual differences in working memory.” In: *Psychological Review* 99.1 (1992), pp. 122–149.

[10] Richard Futrell, Kyle Mahowald, and Edward Gibson. “Large-scale evidence of dependency length minimization in 37 languages.” In: *Proceedings of the National Academy of Sciences* 112.33 (2015), pp. 10336–10341.

[11] Daniel Grodner and Edward Gibson. “Consequences of the serial nature of linguistic input for sentential complexity.” In: *Cognitive Science* 29 (2005), pp. 261–291.

[12] Brian Bartek et al. “In Search of On-line Locality Effects in Sentence Comprehension.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37.5 (2011), pp. 1178–1198.

[13] Roger Levy, Evelina Fedorenko, and Edward Gibson. “The syntactic complexity of Russian relative clauses.” In: *Journal of Memory and Language* 69 (2013), pp. 461–495.

[14] Samar Husain, Shravan Vasishth, and Narayanan Srinivasan. “Strong Expectations Cancel Locality Effects: Evidence from Hindi.” In: *PLoS ONE* 9.7 (2014), pp. 1–14.

[15] Molood Sadat Safavi, Samar Husain, and Shravan Vasishth. “Dependency resolution difficulty increases with distance in Persian separable complex predicates: Implications for expectation and memory-based accounts.” In: *Frontiers in Psychology* 7 (403 2016).

[16] Samar Husain, Shravan Vasishth, and Narayanan Srinivasan. "Integration and prediction difficulty in Hindi sentence comprehension: Evidence from an eye-tracking corpus." In: *Journal of Eye Movement Research* 8(2) (3 2015), pp. 1–12.

[17] Julie A. Van Dyke and Clinton L. Johns. "Memory interference as a determinant of language comprehension." In: *Language and Linguistics Compass* 6.4 (2012), pp. 193–211.

[18] Lena A. Jäger et al. "The subject-relative advantage in Chinese: Evidence for expectation-based processing." In: *Journal of Memory and Language* 79–80 (2015), pp. 97–120.

[19] Shravan Vasishth et al. "Processing Chinese Relative Clauses: Evidence for the subject-relative advantage." In: *PLoS One* 8.10 (2013), e77006.

[20] Fuyun Wu, Elsi Kaiser, and Shravan Vasishth. "Effects of early cues on the processing of Chinese relative clauses: Evidence for experience-based theories." In: *Cognitive Science* (2017).

[21] Shravan Vasishth et al. "Short-term forgetting in sentence comprehension: Crosslinguistic evidence from head-final structures." In: *Language and Cognitive Processes* 25.4 (2010), pp. 533–567.

[22] Stefan L. Frank, Thijs Trompenaars, and Shravan Vasishth. "Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics?" In: *Cognitive Science* 40 (2015), pp. 554–578.

[23] Roger Levy and Frank Keller. "Expectation and locality effects in German verb-final structures." In: *Journal of Memory and Language* 68 (2013), pp. 199–222.

[24] Richard L. Lewis. "An Architecturally-based Theory of Human Sentence Comprehension." Doctoral dissertation. Pittsburgh, PA: Carnegie Mellon University, 1993.

[25] Richard L. Lewis. "Interference in short-term memory: The magical number two (or three) in sentence processing." In: *Journal of Psycholinguistic Research* 25(1) (1996), pp. 93–115.

[26] Julie A. Van Dyke. "Parsing as Working Memory Retrieval: Interference, Decay, and Priming Effects in Long Distance Attachment." Doctoral dissertation. PA: University of Pittsburgh, 2002.

[27] Richard L. Lewis and Shravan Vasishth. "An activation-based model of sentence processing as skilled memory retrieval." In: *Cognitive Science* 29.3 (2005), pp. 375–419.

[28] Julie A. Van Dyke and Richard L. Lewis. "Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities." In: *Journal of Memory and Language* 49 (2003), pp. 285–316.

[29] Julie A. Van Dyke and Brian McElree. "Retrieval interference in sentence comprehension." In: *Journal of Memory and Language* 55.2 (2006), pp. 157–166.

[30] Julie A. Van Dyke. "Interference effects from grammatically unavailable constituents during sentence processing." In: *Journal of Experimental Psychology. Learning, Memory, and Cognition* 33.2 (2007), pp. 407–430.

[31] Julie A. Van Dyke and Brian McElree. "Cue-dependent interference in comprehension." In: *Journal of Memory and Language* 65.3 (2011), pp. 247–263.

[32] Lena A. Jäger, Felix Engelmann, and Shravan Vasishth. "Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis." In: *Journal of Memory and Language* 94 (2017), pp. 316–339.

[33] Richard L. Lewis, Shravan Vasishth, and Julie A. Van Dyke. "Computational principles of working memory in sentence comprehension." In: *Trends in Cognitive Sciences* 10.10 (2006), pp. 447–454.

[34] Brian McElree. "Sentence comprehension is mediated by content-addressable memory structures." In: *Journal of Psycholinguistic Research* 29.2 (2000), pp. 111–123.

[35] Brian McElree. "Accessing recent events." In: *Psychology of Learning and Motivation* 46 (2003). Ed. by B.H. Ross, pp. 155–200.

[36] John R. Anderson et al. "An integrated theory of the mind." In: *Psychological Review* 111.4 (2004), pp. 1036–60.

[37] Bruno Nicenboim and Shravan Vasishth. "Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling." In: *Journal of Memory and Language* 99 (2018), pp. 1–34.

[38] Shravan Vasishth et al. "Processing Polarity: How the ungrammatical intrudes on the grammatical." In: *Cognitive Science* 32.4 (2008).

[39] Daniel Parker and Colin Phillips. "Reflexive attraction in comprehension is selective." In: *Journal of Memory and Language* 94 (2017), pp. 272–290.

[40] Brian W. Dillon et al. "Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence." In: *Journal of Memory and Language* 69 (2013), pp. 85–103.

[41] Bruno Nicenboim et al. "When high-capacity readers slow down and low-capacity readers speed up: Working memory differences in unbounded dependencies." In: *Frontiers in Psychology* 7.280 (2016). ISSN: 1664-1078.

[42] Dave Kush and Colin Phillips. "Local anaphor licensing in an SOV language: Implications for retrieval strategies." In: *Frontiers in Psychology* 5.1252 (2014).

[43] Umesh Patil, Shravan Vasishth, and Richard L. Lewis. "Retrieval interference in syntactic processing: The case of reflexive binding in English." In: *Frontiers in Psychology* 7.329 (2016).

[44] Daniel Parker and Colin Phillips. "Negative polarity illusions and the format of hierarchical encodings in memory." In: *Cognition* 157 (2016), pp. 321–339.

[45] Lena A. Jäger, Felix Engelmann, and Shravan Vasishth. "Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling." In: *Frontiers in Psychology* 6.617 (2015).

[46] Umesh Patil, Shravan Vasishth, and Richard L. Lewis. "Retrieval interference in syntactic processing: The case of reflexive binding in English." In: *Frontiers in Psychology* (2016).

[47] Bruno Nicenboim et al. “Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German.” In: *Cognitive Science* 42 (2018), pp. 1075–1100.

[48] Lena A. Jäger et al. “Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study.” In: *Journal of Memory and Language* (2019). Accepted pending minor revisions.

[49] Marisa Boston et al. “Parallel processing and sentence comprehension difficulty.” In: *Language and Cognitive Processes* 26.3 (2011), pp. 301–349.

[50] Umesh Patil et al. “A computational evaluation of sentence comprehension deficits in aphasia.” In: *Cognitive Science* 40 (2016), pp. 5–50.

[51] Paul Mätzig et al. “A computational investigation of sources of variability in sentence comprehension difficulty in aphasia.” In: *Topics in Cognitive Science* 10.1 (2018). Allen Newell Best Student-Led Paper Award at MathPsych/ICCM 2017, pp. 161–174.

[52] Paula Lissón et al. “Competing models of retrieval in sentence processing: the case of aphasia.” In: 25th Architectures and Mechanisms for Language Processing (AMLaP). Moscow, Russia: Higher School of Economics, 2019.

[53] Felix Engelmann et al. “A framework for modeling the interaction of syntactic processing and eye movement control.” In: *Topics in Cognitive Science* 5.3 (2013), pp. 452–474.

[54] Jakub Dotlačil. “Building an ACT-R Reader for Eye-Tracking Corpus Data.” In: *Topics in Cognitive Science* 10.1 (2018), pp. 144–160.

[55] Fernanda Ferreira, Vittoria Ferraro, and Karl G. D. Bailey. “Good-enough representations in language comprehension.” In: *Current Directions in Psychological Science* 11 (2002), pp. 11–15.

[56] Felix Engelmann. “Toward an integrated model of sentence processing in reading.” Doctoral dissertation. Potsdam, Germany: University of Potsdam, 2016.

[57] Nathan E Rasmussen and William Schuler. “Left-Corner Parsing With Distributed Associative Memory Produces Surprisal and Locality Effects.” In: *Cognitive science* (2017).

[58] Garrett Smith, Julie Franck, and Whitney Tabor. “Semantic features unpack notional plurality in pseudopartitive agreement: A self-organizing approach.” In: *Cognitive Science* 42 (2018), pp. 1043–1074.

[59] Daniel Parker. “Cue Combinatorics in Memory Retrieval for Anaphora.” In: *Cognitive science* 43.3 (2019), e12715.

[60] Christopher Hammerly, Adrian Staub, and Brian Dillon. “The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence.” In: *Cognitive psychology* 110 (2019), pp. 70–104.

[61] John T. Wixted. “Analyzing the empirical course of forgetting.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16.5 (1990), pp. 927–935.

[62] Olga C. Watkins and Michael J. Watkins. "Buildup of proactive inhibition as a cue-overload effect." In: *Journal of Experimental Psychology: Human Learning and Memory* 104.4 (1975), pp. 442–452.

[63] John R. Anderson and Christian Lebiere. *Atomic Components of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1998.

[64] Roger Ratcliff. "A theory of memory retrieval." In: *Psychological Review* 85.2 (1978), pp. 59–108.

[65] Brian McElree, Stephani Foraker, and Lisbeth Dyer. "Memory structures that subserve sentence comprehension." In: *Journal of Memory and Language* 48 (2003), pp. 67–91.

[66] Brian McElree. "Accessing recent events." In: *The Psychology of Learning and Motivation: Advances in Research and Theory*. Ed. by Brian H. Ross. Vol. 46. San Diego, CA: Elsevier, 2006, pp. 155–200.

[67] Matthew Wagers, Ellen F. Lau, and Colin Phillips. "Agreement attraction in comprehension: Representations and processes." In: *Journal of Memory and Language* 61 (2009), pp. 206–237.

[68] Sol Lago et al. "Agreement processes in Spanish comprehension." In: *Journal of Memory and Language* 82 (2015), pp. 133–149.

[69] Matthew A. Tucker, Ali Idrissi, and Diego Almeida. "Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic." In: *Frontiers in Psychology* 6.347 (2015).

[70] Ian Cunnings and Patrick Sturt. "Retrieval interference and sentence interpretation." In: *Journal of Memory and Language* 102 (2018), pp. 16–27.

[71] Serine Avetisyan, Sol Lago, and Shravan Vasishth. "Does case marking affect agreement attraction in comprehension?" In: *Journal of Memory and Language* (2019). Accepted pending revisions.

[72] Brian W. Dillon. "Structured Access in Sentence Comprehension." PhD thesis. College Park, MD: University of Maryland, 2011.

[73] Whitney Tabor, Bruno Galantucci, and Daniel Richardson. "Effects of merely local syntactic coherence on sentence processing." In: *Journal of Memory and Language* 50 (2004), pp. 355–370.

[74] Dario Paape and Shravan Vasishth. "Local coherence and preemptive digging-in effects in German." In: *Language and Speech* 59 (3 2016), pp. 387–403.

[75] David H. Raab. "Statistical facilitation of simple reaction times." In: *Transactions of the New York Academy of Sciences* 24.5 Series II (1962), pp. 574–590.

[76] R. J. Audley and A. R. Pike. "Some alternative stochastic models of choice." In: *British Journal of Mathematical and Statistical Psychology* 18.2 (1965), pp. 207–225. ISSN: 2044-8317.

[77] Douglas Vickers. "Evidence for an Accumulator Model of Psychophysical Discrimination." In: *Ergonomics* 13.1 (Jan. 1970), pp. 37–58. ISSN: 1366-5847.

[78] Marius Usher and James L McClelland. “The time course of perceptual choice: The leaky, competing accumulator model.” In: *Psychological Review* 108.3 (2001), p. 550.

[79] Sylvia Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.

[80] Shravan Vasishth et al. “Modelling dependency completion in sentence comprehension as a Bayesian hierarchical mixture process: A case study involving Chinese relative clauses.” In: *Proceedings of the Cognitive Science Conference*. London, UK, 2017.

[81] Bob Carpenter et al. “Stan: A probabilistic programming language.” In: *Journal of Statistical Software* 76.1 (2017).

[82] Aki Vehtari, Janne Ojanen, et al. “A survey of Bayesian predictive methods for model assessment, selection and comparison.” In: *Statistics Surveys* 6 (2012), pp. 142–228.

[83] Aki Vehtari, Andrew Gelman, and Jonah Gabry. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.” In: *Statistics and Computing* 27.5 (2017), pp. 1413–1432.

[84] Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. “Assessing the ability of LSTMs to learn syntax-sensitive dependencies.” In: *Transactions of the Association for Computational Linguistics* 4 (2016), pp. 521–535.

[85] David Caplan, Jennifer Michaud, and Rebecca Hufford. “Mechanisms underlying syntactic comprehension deficits in vascular aphasia: New evidence from self-paced listening.” In: *Cognitive Neuropsychology* 32.5 (2015), pp. 283–313.

[86] Sandra Hanne et al. “Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction.” In: *Journal of Neurolinguistics* 34 (2015), pp. 83–111.

[87] Sandra Hanne, Frank Burchert, and Shravan Vasishth. “On the nature of the subject-object asymmetry in wh-question comprehension in aphasia: Evidence from eye-tracking.” In: *Aphasiology* (2015).

[88] Felix Engelmann, Lena A. Jäger, and Shravan Vasishth. “The effect of prominence and cue association in retrieval processes: A computational account.” Manuscript submitted to Cognitive Science. 2019.

[89] Seth Roberts and Harold Pashler. “How persuasive is a good fit? A comment on theory testing.” In: *Psychological Review* 107.2 (Apr. 2000), pp. 358–367.

[90] Jelte M. Wicherts et al. “The poor availability of psychological research data for reanalysis.” In: *American Psychologist* 61.7 (2006), p. 726.

[91] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.

[92] Jacob Cohen. “The statistical power of abnormal-social psychological research: A review.” In: *The Journal of Abnormal and Social Psychology* 65.3 (1962), p. 145.

[93] Open Science Collaboration et al. “Estimating the reproducibility of psychological science.” In: *Science* 349.6251 (2015), aac4716.

[94] Daniel J. Benjamin et al. “Redefine statistical significance.” In: *Nature Human Behaviour* 2.1 (2018), p. 6.

[95] Blakeley B. McShane et al. “Abandon statistical significance.” In: *The American Statistician* 73.sup1 (2019), pp. 235–245.

[96] Shravan Vasishth et al. “The statistical significance filter leads to overoptimistic expectations of replicability.” In: *Journal of Memory and Language* 103 (2018), pp. 151–175.

[97] Andrew Gelman and John Carlin. “Beyond Power Calculations Assessing Type S (Sign) and Type M (Magnitude) Errors.” In: *Perspectives on Psychological Science* 9.6 (2014), pp. 641–651.

[98] Andrew Gelman et al. *Bayesian Data Analysis*. Third. Chapman and Hall/CRC, 2014.

[99] Daniel J. Schad, Michael Betancourt, and Shravan Vasishth. “Towards a principled Bayesian workflow: A tutorial for cognitive science.” Unpublished manuscript. 2019.

[100] David Spiegelhalter and Michael Blastland. *The Norm chronicles: Stories and numbers about danger*. Profile Books, 2013.

[101] Colin Phillips, Matthew Wagers, and Ellen F. Lau. “Grammatical illusions and selective fallibility in real-time language comprehension.” In: *Experiments at the Interfaces*. Ed. by Jeff Runner. Vol. 37. Syntax and Semantics. Bingley, UK: Emerald Group Publishing Limited, 2011, pp. 147–180.

[102] Ronald L. Wasserstein and Nicole A. Lazar. “The ASA’s Statement on p-Values: Context, Process, and Purpose.” In: *The American Statistician* 70.2 (2016), pp. 129–133.

[103] Daniel J. Schad and Shravan Vasishth. “The posterior probability of a null hypothesis given a statistically significant result.” Submitted. 2019.

[104] Evan Kidd, Seamus Donnelly, and Morten H. Christiansen. “Individual differences in language acquisition and processing.” In: *Trends in Cognitive Sciences* 22.2 (2018), pp. 154–169.

[105] Thomas A. Farmer, Jennifer B Misyak, and Morten H. Christiansen. “Individual differences in sentence processing.” In: *Cambridge handbook of psycholinguistics* (2012), pp. 353–364.

[106] Julia M. Haaf and Jeffrey N. Rouder. “Some do and some don’t? Accounting for variability of individual difference structures.” In: *Psychonomic Bulletin and Review* (2018), pp. 1–18.

[107] Donald R. Williams, Jeffrey N. Rouder, and Phillipe Rast. “Beneath the Surface: Unearthing Within-Person Variability and Mean Relations with Bayesian Mixed Models.” Unpublished draft. 2019.

[108] Justine B. Wells et al. “Experience and Sentence Comprehension: Statistical Learning, Working Memory, and Individual Differences.” In: *Cognitive Psychology* (8888).

[109] Pavel Logačev and Shravan Vasishth. “Understanding underspecification: A comparison of two computational implementations.” In: *Quarterly Journal of Experimental Psychology* 69.5 (2016), pp. 996–1012.

[110] David Caplan. “Resource reduction accounts of syntactically based comprehension disorders.” In: *Perspectives on agrammatism*. Ed. by Cynthia K. Thompson and R. Bastiannse. Psychology Press, 2012, pp. 34–48.

[111] Petra Burkhardt, Maria Mercedes Piñango, and Keng Wong. “The role of the anterior left hemisphere in real-time sentence comprehension: Evidence from split intransitivity.” In: *Brain and Language* 86.1 (2003), pp. 9–22.

[112] Paul-Christian Bürkner. “brms: An R Package for Bayesian Multilevel Models Using Stan.” In: *Journal of Statistical Software* 80.1 (2017), pp. 1–28.

Box 1: Inhibitory and facilitatory interference

Inhibitory interference: A schematic illustration of inhibitory interference is shown in Figure I; here, retrieval cues can match multiple chunks in memory (1a), leading to inhibitory interference. An example from [40] illustrates the retrieval process. When the parsing system attempts to complete the dependency between the verb phrase *was* and its grammatical subject, *bodybuilder*, a search is triggered in memory for a singular noun phrase that is the subject of the main clause. If another singular noun phrase, e.g., *trainer*, intervenes between the subject noun and the verb, this is predicted to lead to inhibitory interference.

Facilitatory interference: A schematic illustration of facilitatory interference is shown in Figure I; in (2a), each of two retrieval cues match two different chunks in memory, leading to faster reading time at the auxiliary verb *were*. Cf. the baseline condition (2b), where one of the retrieval cues matches only the target noun. The reason for the facilitation is the race process that is inherent in the cue-based retrieval mechanism in ACT-R (see Figure 2).

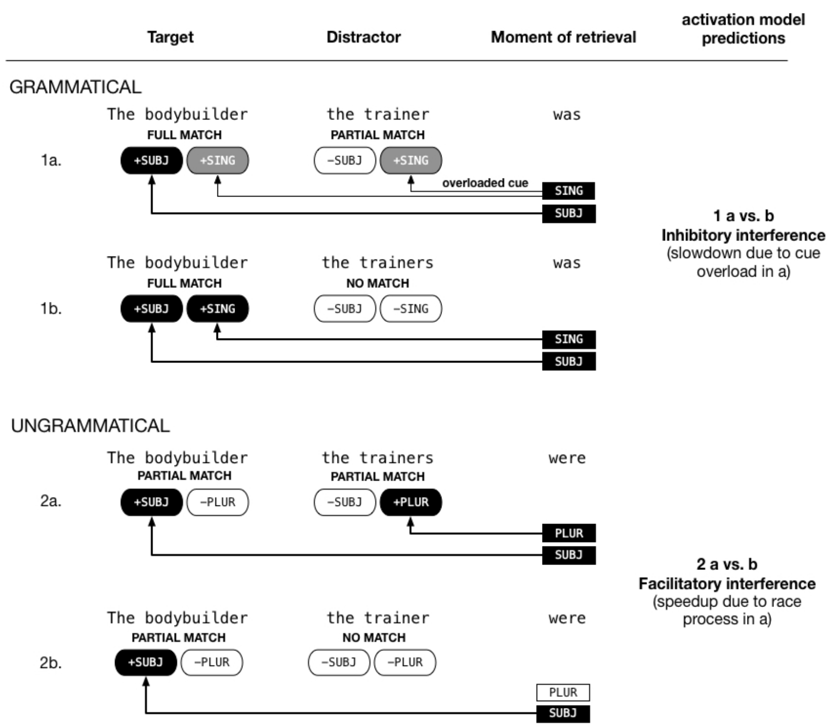

Fig. I. Predictions of the activation model for grammatical and ungrammatical configurations. The model predicts that grammatical conditions show inhibitory interference effects, and ungrammatical conditions show facilitatory interference effects. The quantitative predictions in each case are shown in Figure 1.

Box 2: Comparing the activation and direct-access model The direct-access model can be defined as a finite mixture model as follows. Let y be the reading time in milliseconds, and β the mean time in log milliseconds taken for a successful retrieval, with standard deviation σ. Such a successful retrieval happens with probability p. Retrieval is assumed to fail with probability $(1-p)$, and the extra cost of re-attempting and successfully carrying out retrieval is δ log ms. For the full, hierarchically specified model, see [37].

$$y \sim \begin{cases} LogNormal(\beta, \sigma^2), & \text{retrieval succeeds, probability } p \\ LogNormal(\beta + \delta, \sigma^2), & \text{retrieval fails initially, probability } 1 - p \end{cases} \tag{1}$$

We can now determine whether the observed data are underlyingly coming from a two-component mixture or from a LogNormal race, and whether a mixture distribution yields better predictions with respect to the data. Figure II shows a comparison of the relative predictive fits from the hierarchical Bayesian models implementing the activation model as a race of accumulators, and McElree's direct-access model as a finite mixture process. The violin plots show posterior predictive distributions from the model; their width represents the density of the predicted mean reading times. The black circles show the empirically observed mean reading times. The four types of reading times refer to four different kinds of question responses that the participants gave in a self-paced reading task on inhibitory interference [47]. The activation model overestimates the reading times in the incorrect responses, compared to the direct-access model. Although not shown here, this overestimation is due to the activation model assuming a single variance component for both correct and incorrect responses. When that assumption is relaxed, both models show similar predictive accuracy [37].

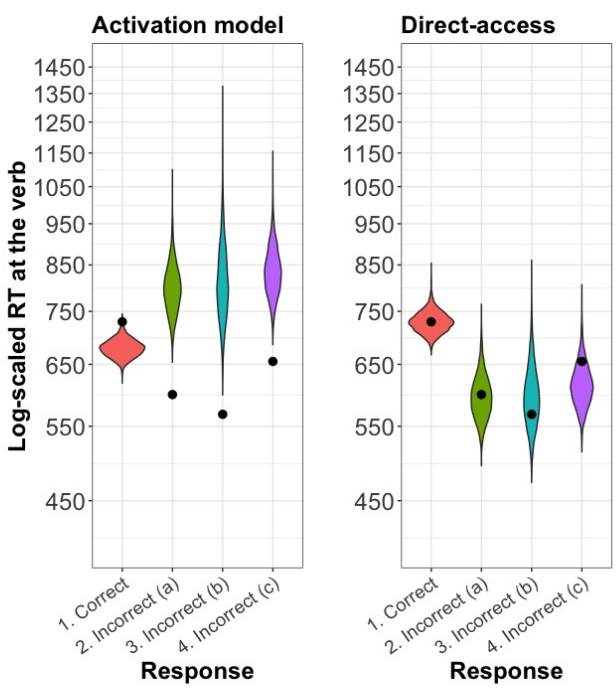

Fig. II. A comparison of observed sample means with the posterior predictive distributions of the activation model and the direct-access model.

Box 3: Modeling impairments in sentence comprehension in aphasia
Using the activation model, [51] investigated three theories processing deficits in aphasia. *Intermittent deficiency* [85]—occasional breakdowns in parsing—can be implemented within the activation model as arising from increased noise in activation; the higher the noise, the greater the fluctuation in activation and the greater the amount of processing difficulty on average, and the higher the probability of parsing failure. *Resource reduction* [110] can be modeled by reducing the total amount of activation (called goal activation in ACT-R) that can be allocated to a chunk that matches a particular set of retrieval cues—the higher the allocation of the activation, the greater the amount of activation (and therefore resources) available to the comprehender. *Slowed processing* [111] can be implemented as increased default processing time for each parsing step in the model; slowed parsing steps will lead to reductions in activation and therefore greater processing difficulty. Using self-paced listening data on subjects vs. object relative clauses from 56 individuals with aphasia (IWAs) and 46 matched controls [85], [51] estimated, for each participant separately, the numerical parameters corresponding to each of the above theoretical proposals. Figure III shows the distribution of the best parameter estimates for each participant in subject vs. object relatives. The plots show that compared to controls, IWAs tend to have lower goal activations, higher noise, and slower default action time, suggesting that IWA experience these deficits to a greater extent than controls, but each of the deficits is present in each IWA to differing degrees. The fact that some controls also show such patterns suggests that the deficits along these three dimensions may all in principle be candidate theories for characterizing processing difficulty both in IWAs and controls.

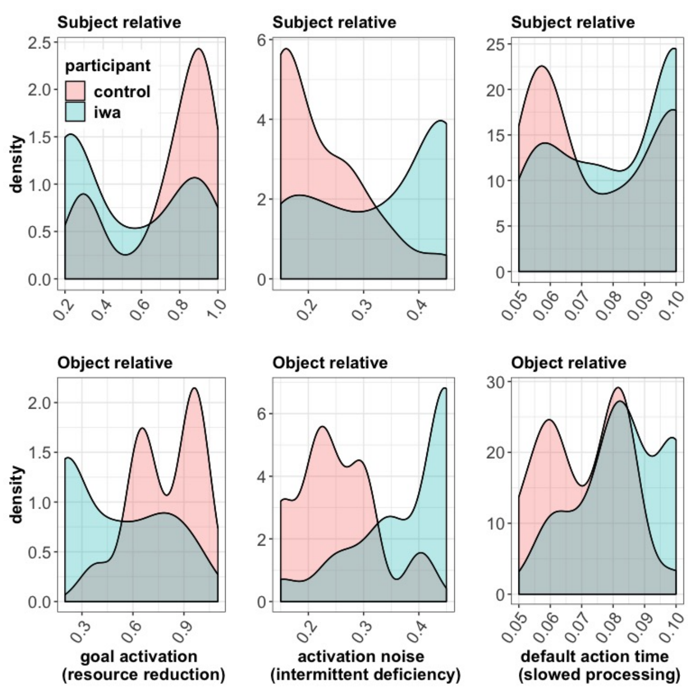

Fig. III. Shown are the distributions of the best parameter estimates in the activation model for each individual participant (individuals with aphasia or controls).

Outstanding questions

- Several competing computational models of sentence processing are now available, but their relative predictive performance needs to be systematically tested against existing benchmark data.
- Appropriately powered, large-sample benchmark data are needed from a broad spectrum of syntactic phenomena and languages in order to evaluate quantitative model predictions.
- A potentially rich and until now underdeveloped area for research is using models of unimpaired processing to attempt to explain sources of variability in comprehension difficulty in impaired populations.
- Quantitative models should aim to explain both group-level effects as well as the (often systematic) variability seen in individual participants.

These important open area provide a rich set of opportunities for both empirical researchers and computational modelers.