Speed-accuracy trade-off modelling and its interface with experimental syntax

Book or Report Section

Accepted Version

Foraker, S., Cunnings, I. ORCID: https://orcid.org/0000-0002-5318-0186 and Martin, A. E. (2023) Speed-accuracy trade-off modelling and its interface with experimental syntax. In: Sprouse, J. (ed.) The Oxford Handbook of Experimental Syntax. Oxford Handbooks. Oxford University Press, Oxford, pp. 363-392. ISBN 9780198797722 doi: 10.1093/oxfordhb/9780198797722.013.12 Available at https://centaur.reading.ac.uk/111284/

To link to this article DOI: http://dx.doi.org/10.1093/oxfordhb/9780198797722.013.12

Publisher: Oxford University Press

Chapter 16

Speed-Accuracy Tradeoff Modeling and its Interface with Experimental Syntax

Stephani Foraker[1], Ian Cunnings[2], & Andrea E. Martin[3]

[1] Psychology Department, State University of New York College at Buffalo

[2] School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

[3] Psychology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

In this chapter, we review key insights gained by using the speed-accuracy tradeoff (SAT) technique to address psycholinguistic and linguistic issues. SAT evidence has been instrumental in integrating sophisticated memory models into psycholinguistic theory, and bears on several linguistic issues in experimental syntax. We explain how SAT can provide clear evidence about time course of processing that is unconfounded by accuracy or probability of interpretation over trials, and in so doing, can fruitfully inform debates about processing and representation.

Many advances in linguistic theory have been made via acceptability judgements (Sprouse, Schütze, and Almeida 2013). Yet, linguistic judgements, like any other kind of judgement, are inherently susceptible to a speed-accuracy tradeoff – a phenomenon where the probability of making a particular judgement can change as information accumulates over time. For example, it is well attested in agreement attraction that ungrammatical sentences, such as 'The key to the cabinets were rusty', may be perceived as grammatical when a reader is under time pressure, even though native language users may be more likely to reject this type of sentence as ungrammatical when given enough time (e.g., Parker, 2015; Phillips, Wagers, and Lau 2011; Wagers, Lau, and Phillips 2009). This indicates that perception of sentence acceptability, or indeed, perception of sentence plausibility or the availability of a particular sentence interpretation, interacts with the time taken to make a particular response. The SAT technique provides an explicit way of assessing this inherent tradeoff between speed and accuracy in linguistic judgements as they unfold over time.

One well-motivated reason for why linguistic judgements are susceptible to a speed-accuracy tradeoff is that making such judgements involves accessing information from memory, particularly when dependent elements are separated by intervening material. In cases of agreement attraction, such as 'The key to the cabinets were rusty', correctly rejecting the sentence as ungrammatical requires accessing the sentence subject ('the key') from memory at the verb ('were'), and assessing the dependency between these elements as ungrammatical, due to the number mismatch. Attraction errors occur when the intervening, linearly closer, constituent ('the cabinets') is bound[1] with the verb based on the matching number features, instead of querying memory to retrieve the syntactic, singular subject. That is, dependency creation requires accessing the correct item in memory.

---

[1] We use the terms 'binding' and 'bound' as used in the memory literature to refer to the mechanism by which information in memory is integrated together (e.g., Cohen and Eichenbaum 1993; Hagoort 2003; James 1918), rather than the more specific use of these terms in the linguistics literature to refer to specific cases of anaphoric coreference in the presence of c-command (e.g. Chomsky 1981).

16.1 SAT unconfounds quality of information from time-course of processing

Much of our current understanding of language processing comes from studies using measures of processing time to test hypotheses about the nature of the representations and processes deployed in real-time language production and comprehension. Common timing measures include reaction times for judgements about an expression, reading time or eye-movement measures while reading an expression, or, as in the visual world paradigm, the time at which eye-movements are launched to visual objects in a display as a related spoken utterance is interpreted.

In many, if not most, applications of timing measures, researchers compare an expression containing a property that is hypothesized to tax comprehension operations with a minimally contrastive control expression, with the prediction that the former will take longer to read than the latter. A positive finding provides support for the researcher's hypothesis and evidence against any model that is not able to draw a principled distinction between the two types of expressions. Although untimed measures, such as percent accuracy or acceptability ratings, might reveal a comparable effect, timing measures are often preferred, as they are generally considered more sensitive than other behavioral measures, particularly for highly accurate response measures (Sternberg 1966). Note that in these types of applications, timing data are used simply as an ordinal measure – as a means to verify that one type of expression is indeed more taxing to process than another.

Following seminal work by Sternberg (1969), which reintroduced and extended earlier work by Donders (1868/1969), timing measures have been used to address a host of questions concerning the nature and organization of mental architectures. Crucially, the key assumption in such applications is that differences in timing measures scale to differences in real-time mental operations, functioning as interval or ratio measures. For situations when this assumption holds, it licenses the use of timing measures to investigate components that are essential precursors to developing fully articulated models of language processing, such as when certain processes are operative, how component operations within a complex skill are organized (e.g., in serial or parallel), or whether one source of information by-passes or suppresses the use of another.

To what degree can we be certain that differences in timing measures scale to differences in real-time mental operations? As noted previously, with very few exceptions, most timing measures are sensitive to tradeoffs between speed and accuracy (Wickelgren 1977). In language comprehension tasks comprehenders have flexibility over the depth to which they process an expression in a given context, and (if required) when and how to execute an overt response. Consequently, differences in timing measures can reflect differences in subjective criteria rather than just intrinsic differences in the speed of processing. Language scientists often try to control for or rule out criterion shifts by assessing measures of processing accuracy, either direct or indirect, to supplement basic timing measures. However, the approaches are valid only if accuracy level is measured *at the same point in time* that the timing measure is collected.

A second, more formidable concern is that common timing measures are not pure measures of underlying processing speed. Rather, they are sensitive to factors that affect comprehension accuracy, particularly the quality and availability of information required to construct a meaningful interpretation. Furthermore, although language comprehension engages a set of highly overlearned, largely automatic cognitive processes, it is not without error. For example, two types of expressions may differ in reading time or acceptability judgment latency if (a) the quality of the resulting interpretations substantially vary (e.g., their acceptability, plausibility, specificity, etc.), or (b) errors in key operations are more likely in one than the other (e.g., failure to retrieve essential information, misanalyses of grammatical relations). Such factors that affect quality or availability of information lead to differences in interpretation accuracy, which can vary *independently* of the time taken to compute that interpretation. As a consequence, researchers cannot straightforwardly interpret a difference in most commonly used timing measures purely as an underlying difference in processing speed. The SAT method, then, provides separable measures of speed and accuracy by modeling how comprehension accuracy develops over processing time.

To help illustrate this central problem with common timing measures, consider sentences (1)-(4).

(1) The doctor realized the boy yelled.

(2) This is the boy [who the doctor realized] yelled.

(3) This is the boy [who the doctor [who calmed the mother] realized] yelled.

(4) This is the boy [who the doctor [who ordered a blood test] realized] yelled.

Adding material between the dependent elements of ‘boy’ and ‘yelled’ typically introduces differences that affect the accuracy of the resulting interpretations due to differing availability or quality of information necessary to create and resolve the dependency. For example, the interpolated clause(s), shown in square brackets, may alter expectations about upcoming information. They may also increase the likelihood of misparsing the expression, as the interpolated material may introduce alternative attachment sites with some degree of local coherence that the reader does not recover from on some proportion of trials (e.g., in (3), did the mother yell or the boy yell?). Even if expectations are held constant and the potential for misparsing is minimized, as we will discuss, many studies now leave little doubt that interpolated material can adversely affect sentence comprehension. All of these factors can affect the quality and accuracy of interpretation, without necessarily affecting the time course of processing.

The predictions that are crucial for evaluating many fundamental hypotheses about language processing, particularly those that address basic architectural issues, often concern the respective speed of processing for different types of expressions. Time course measures are critical for answering questions such as whether the interpretation of one type of expression increases the complexity of a particular operation, whether it recruits an altogether different type of operation, or whether it requires more operations than another (e.g., Bott, Bailey, and Grodner 2012; Bott, Rees, and Frisson 2016; McElree and Griffith 1995, 1998; McElree and Nordlie 1999; McElree, Pylkkänen, Pickering, and Traxler 2006). Time course measures also provide the primary means of investigating general architectural issues, such as whether there are contingencies in the organization of component operations, with some operations having temporal priority over others, versus organized in an interactive fashion (Bornkessel, McElree, Schlesewsky, and Friederici 2004; Martin and McElree 2018; McElree 1993). Hence, the great value of SAT measures, in contrast to other timing measures, is that they unconfound the quality or availability of information from the time course of processing that information.

16.2 The SAT methodology

In most situations when timing measures are collected, it is up to the participant to strike a reasonable balance between the competing demands of accuracy and speed. Typically, a speed-accuracy tradeoff allows the probability of making a correct decision to increase as information accumulates over time. In the SAT procedure, this tradeoff is for the most part controlled by having participants respond at set time points. As quickly as possible after a signal – usually an auditory tone for a reading task – the participant makes a response. Signal time points are chosen to capture the full span of processing from before the beginning of a critical word or region until after the end, providing a window onto processing as it unfolds over time. The signals can be distributed either across trials with one deadline per trial (single-response SAT, for illustration, see McElree 2000), or they can all occur on each trial (multiple-response SAT, for illustration, see McElree 1993). A critical aspect of both SAT variants – one that distinguishes the SAT from all other alternative measures of processing speed (see Wickelgren 1977) – is that participants do not themselves decide when they want to respond. By constraining participants to respond at both early and late times, researchers control potential speed-accuracy tradeoffs and can thereby obtain unbiased measures of accuracy and rate of processing at each time point.

In SAT investigations of language comprehension, researchers have required participants to discriminate acceptable from unacceptable expressions, with sets that include yoked acceptable and unacceptable conditions. For example, to measure the speed and accuracy of resolving the long-distance dependency between 'yelled' and 'the boy' in (2), the participant judges whether the sentence 'This is the boy who the doctor realized yelled' is acceptable or not, and on another trial judges whether an unacceptable counterpart such as 'This is the boy who the doctor realized tore*' is acceptable or not. The final verb, 'yelled' vs. 'tore', is the locus of the critical binding operation at question here, and hence, the probes signaling the participant to make an acceptability judgment need to begin just before that critical point of the sentence. Furthermore, it is of fundamental importance that the source of unacceptability clearly taps the particular issue being investigated, and is not confounded with other sources of unacceptability. If the locus of the unacceptability is not carefully and well-chosen to target the dependency under investigation, interpretation of the data will be murky at best. In sentences (2)-(4), the syntactic structures are unacceptable sentences only when comprehenders attempt to bind the clefted NP to the incompatible verb (‘the boy tore’). Requiring participants to discriminate acceptable from unacceptable expressions obliges them to determine the acceptability of the final verb and clefted subject to obtain above chance performance.

A strength of SAT methodology is that the judgement data are fit individually for each participant, and typically for each item, and then patterns across participants and items are evaluated for consistency. One key consideration is that to calculate stable SAT functions many observations for each participant and item are necessary. Therefore, in most applications of SAT, each participant is presented with all versions of an item. For examples (1)-(4), representing four conditions, participants would encounter eight versions of an item; four acceptable, four unacceptable. The eight versions would be counterbalanced across sessions on different days and shown in a different order to each participant, but every participant would judge all eight versions of each item. This is in contrast to contemporary psycholinguistic experiment designs, where different conditions of an item are counterbalanced across participants.

SAT results have been criticized on these grounds, since repeated exposure may introduce different routines or strategies that do not reflect typical processing. However, several investigations have revealed converging evidence from eyetracking of reading (Foraker and McElree 2007; Martin and McElree 2008, 2011; Van Dyke and McElree 2011), as well as judgement latency (McElree 1993; McElree and Griffith 1995, 1998) and ERP measures (Bornkessel et al. 2004), reducing concerns about the generalizability and construct validity of SAT evidence. Additionally, including astute control conditions and/or additional experiments to rule out alternative explanations for the observed results remain a critical element of effective experimental design (e.g., see Martin and McElree 2008; McElree and Griffith 1998 for discussion).

To calculate SAT functions, responses are typically corrected for response biases by transforming percent correct into d-prime (*d'*) for each time point. One reader may be more liberal overall, registering more acceptable judgements for the materials, while another may be more conservative, with fewer acceptable judgements; *d'* provides a way to standardize across this variation. For binary yes-no judgements, an equal-variance Gaussian *d'* is typically used, which is the *z*-transform of the hit rate minus the *z*-transform of the false alarm rate, $d' = z(P(\text{"yes"}|\text{acceptable})) - z(P(\text{"yes"}|\text{unacceptable}))$[2]. A hit is when an acceptable sentence is correctly interpreted as acceptable (‘boy’ is bound with ‘yelled’), while a false alarm is when an unacceptable sentence is incorrectly interpreted as acceptable (‘boy’ is bound with ‘tore’).

Hypothetical data shown in Figure 16.1 illustrate accuracy (black circles) at different response signal time points for one condition. A curve representing the growth of response accuracy is also shown (solid line), fit with an exponential approach to a limit, plotting accuracy (*d'*) as a function of processing time (*t*): $d' = \lambda (1 - e^{-\beta (t - \delta)})$ for $t > \delta$, otherwise 0. A *d'* of zero equals chance performance, with scores approaching 4 reflecting nearly perfect performance.

[2] Pseudo-*d'* can also be computed to examine incorrect responses to the target experimental sentences; see McElree (1998), McElree and Dosher (1989).

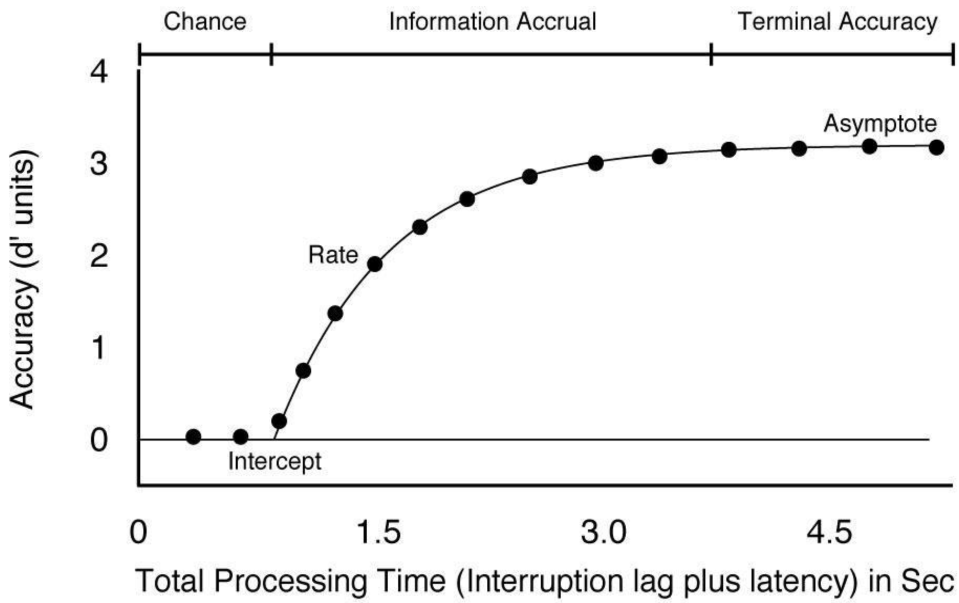

Figure 16.1 An SAT function for one condition, illustrating the three phases of processing.

SAT functions show an initial period of chance performance followed by a monotonically increasing function, culminating in the final asymptotic level. The three parameters of the exponential, λ, β, and δ, are used to estimate how conditions vary in the three phases of processing illustrated in Figure 16.1. The parameter λ represents the *asymptote* of the function, and it provides an estimate of the highest level of discrimination reached with maximal processing time. The parameters δ and β provide joint measures of the speed of processing – also referred to as the time course dynamics – indexing how quickly accuracy accrues to its asymptotic level. The parameter δ estimates the *intercept* of the function, which is when accuracy departs from chance level, and provides an estimate of the point in time at which comprehenders first show sensitivity to the information necessary to discriminate acceptable from unacceptable sentences. The parameter β estimates the slope, or *rate*, at which accuracy grows from chance to asymptote.

Determining how experimental conditions impact the shape of their corresponding SAT function requires a hierarchical model-testing scheme in which different combinations of λ, β, and δ are competitively applied to the SAT functions. First, in order to obtain a robust estimate of asymptote, averaging the final 2-3 *d'* points into a single bin can assure that a larger sample is the basis of the asymptote estimates, in turn allowing for more stable estimates of the dynamics parameters. To find the best fitting set of parameters for different conditions, the number of asymptotes, intercepts, and rates are systematically varied from a null model (one asymptote, rate, and intercept, 1λ-1β-1δ, for all data points from all conditions), through hypothesis-driven combinations, like 4λ-1β-2δ, for example, to a fully saturated model, such as 4λ-4β-4δ for four conditions.

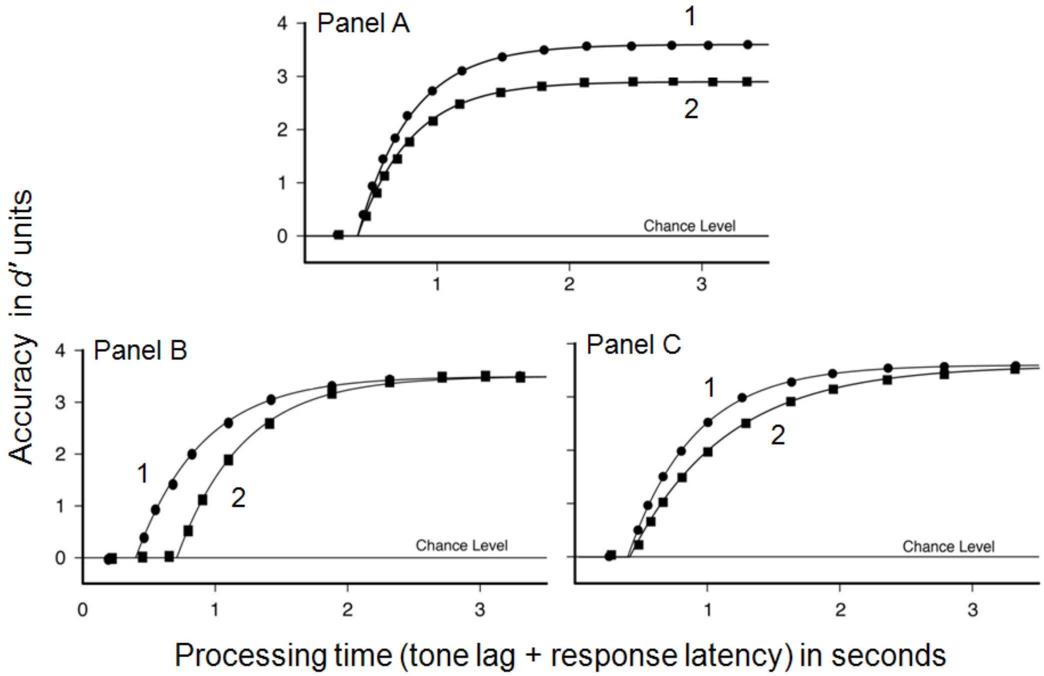

Figure 16.2 Idealized differences in the three phases of the SAT functions for two conditions.

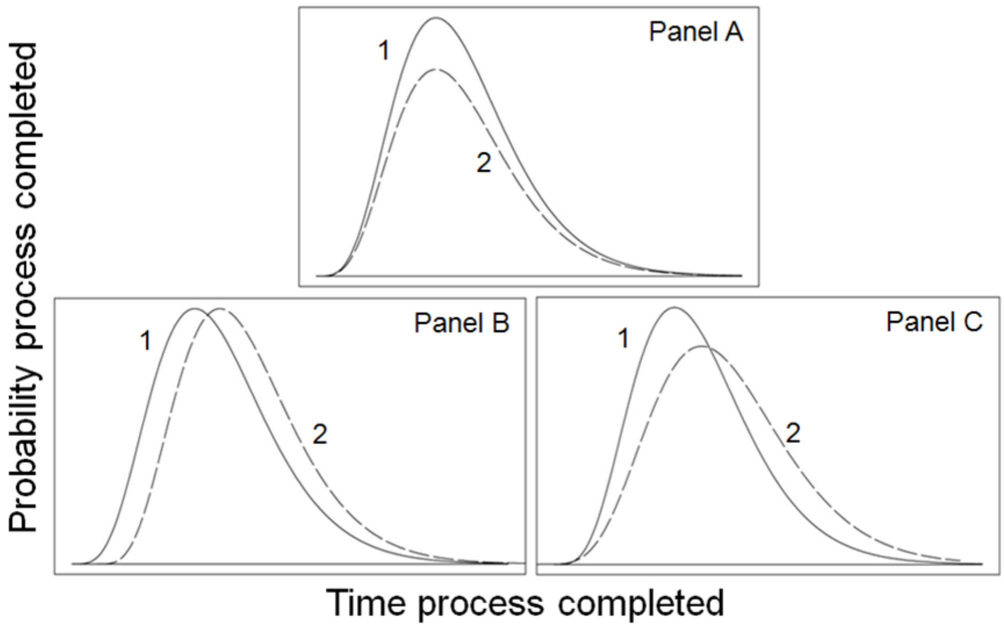

Figure 16.3 Idealized differences in the finishing time distributions corresponding to the SAT differences shown in Figure 16.2.

Panel A of Figure 16.2 illustrates a case where two conditions are associated with the same intercept and rate but differ in asymptotic accuracy, with Condition 2 having a lower asymptote. In terms of the underlying distribution of finishing times for the judgement at hand, illustrated in Panel A of Figure 16.3, this asymptote pattern can arise if fewer processes in one trial and/or across trials successfully complete in Condition 2 than Condition 1. In this illustration, those that do successfully complete do so with a comparable distribution of times, and the overlapping finishing times result in equivalent intercepts and proportional slopes (rate).

Panels B in Figures 16.2 and 16.3 illustrate a case where two conditions are associated with the same asymptote and rate but different intercepts, with Condition 2 having a delayed intercept. Panels C in Figures 16.2 and 16.3 illustrate the case where two conditions are associated with the same asymptote and intercept but different rates, with Condition 2 having a disproportionately slower approach to asymptote. These two patterns reflect differences in the speed of processing only. Differences in SAT intercept correspond to differences in the minimum of the finishing time distributions, whereas differences in SAT rate correspond to differences in the variances of the distributions. For example, if one condition consistently requires additional computational operations applied in a serial or cascading fashion, then relative to a condition with fewer operations, the finishing time distribution will be shifted toward longer times, manifesting as different intercepts. On the other hand, if one condition entails a relaunched operation on some subset or proportion of trials, the distribution will be more positively skewed, leading to a disproportionately slower rate. To wit, reanalysis may be attempted on misparsed trials, or additional queries of memory to retrieve lower quality or less frequent information may be needed (see also McElree and Dosher 1989,1993; Reed, 1976). A lower asymptote for the condition with the slower rate is furthermore consistent with this kind of explanation.

The best fitting model is chosen by a combination of criteria. In most published applications, models have been fit with a least-squares error criterion (Chandler 1969; Reed 1976), with the quality of the fit assessed by goodness-of-fit statistics, such as adjusted-R^2 (Judd and McClelland 1989). Model fits are performed for the averaged data for expository purposes, but it is essential to model each participant’s data separately (and by items), allowing evaluation of the consistency of parameter estimates across participants (and items), and inferential tests of significance.

The most crucial aspects of model fitting are that (a) only differences that exist in observed *d'* should be posited in the models, and (b) such patterns should be largely evident across individual participants and/or items. That is, statistical tests on *d'* across participants and/or items should support a reliable difference between conditions before any difference is posited in the asymptote parameter (λ). Otherwise, the estimates of the SAT dynamics parameters (β, δ) will be biased to account for a difference that does not exist in the *d′* data. At the core of the hierarchical model testing procedure is allocating a minimum number of parameters to best account for the variance in how *d′* accrues over time. As such, the first model fit is a 1λ-1β-1δ model which fits all conditions with the same three parameter values. Even if a difference in *d′* by condition already exists, starting with 1λ-1β-1δ establishes the baseline adjusted-R^2. Next, if there is an observable difference in *d′* between two conditions, the next models to test are 2λ-1β-1δ, 2λ-2β-1δ, 2λ-1β-2δ, and 2λ-2β-2δ. Those four are evaluated for the best fitting model, assessing which SAT parameter estimates differ systematically and reliably across participants (and items) using inferential statistics. The best fitting model will typically have the highest adjusted-R^2 across participants and usually for the average data, but we note that, in our experience, adjusted-R^2 alone is not diagnostic to the best fitting model – some models result in adjusted-R^2 that differ on such small orders of magnitude that it is difficult to assign meaning to that difference. It is therefore more crucial that any differences in parameter values – whether asymptote, rate, or intercept – are reliable across participants (and items). These two points require that empirical *d′* and parameter estimates are reported for individual participants.

If evidence for a difference in SAT dynamics parameters is found, one way to guard against parameter tradeoff – the phenomenon where variance in one parameter is allocated to another parameter – is to perform fixed parameter fits, where, for example, the asymptote parameter (λ) is fixed to the value of *d′* (based on averaging over the last 2-3 time points, as noted above) and not allowed to vary. This will force the model to assign any remaining applicable variance to the other parameters, rather than erroneously account for that variance by modulating the asymptote as well. In the case of a veridical difference in processing speed, reliable differences between conditions across participants should appear even when the asymptotes are not allowed to vary from the *d'* values. A similar approach can be taken when trying to evaluate the relationship between rate (β) and intercept (δ). Intercept can be inferred from the first time lag where *d'* departs from chance; either fixing intercept to this value, or only fitting a single dynamics parameter at a time can be employed to check for parameter tradeoff between rate (β) and intercept (δ).

Finally, the best fit model should be interpreted in light of the competing explanations, accounts, or theories being tested. Significant differences in asymptote along with null results for intercept or rate might support one explanation, while differences in both asymptote and rate are more consistent with another, and so on. What is of interest is in which parameter(s) the differences emerge and for which conditions, and having principled predictions for the presence or absence of such differences. Additionally, replication across SAT experiments and converging evidence from other measures and methods can constrain and solidify inferences.

16.3 Memory operations are fundamental to language processing

Memory-based operations, such as encoding, storage, and retrieval, have long been acknowledged as important factors in language production and comprehension, and in constraining linguistic theory. From the early days of psycholinguistics, limits on center-embedding or other aspects of language complexity illustrated that memory constraints can interact with linguistic content and sentence acceptability (e.g., Miller and Chomsky 1963). Attempts to explain why memory limitations may determine the upper bound on our ability to interpret complex sentence structures have typically appealed to working memory capacity limits, usually focusing on memory storage capacity. Perhaps the most well-known theory of this type is Daneman and Carpenter's (1980) capacity-based theory of sentence comprehension, which explains difficulty in sentence processing in terms of the amount of information that an individual must hold in memory at one time. Consider again examples (2-4) above. Here, successful comprehension requires encoding a representation of 'the boy' when it is first encountered, storing it in working memory whilst the following constituents are processed (and themselves encoded in working memory), and then retrieving it from memory at the verb 'yelled'. Similarly, locality effects have been explained in terms of working memory load, where progressively greater processing difficulty occurs with additional unresolved dependencies, due to maintaining more items in working memory, as well as increasing the distance between elements being integrated (Gibson 2000; Grodner and Gibson 2005; Warren and Gibson 2002).

However, a substantial body of research shows that memory-based restrictions on language are best described not in terms of the amount of information that needs to be held in limited-capacity working memory at one time, but rather in terms of the *retrievability* of information in memory, based on its content and quality (see meta-analysis Jäger, Engelmann, and Vasishth 2017; reviews: Foraker and McElree 2011; Van Dyke and Johns 2012; Parker, Shvartsman, and Van Dyke 2017; also: Gordon, Hendrik, and Johnson 2001, 2004; Gordon, Hendrik, and Levine 2002; Martin 2016; Martin, Nieuwland, and Carreiras 2012, 2014; Van Dyke 2007; Van Dyke and Lewis 2003; Van Dyke and McElree 2006, 2011; Van Dyke, Johns, and Kukona, 2014; Vasishth, Brüssow, Lewis, and Drenhaus 2008). Importantly, this approach suggests a fundamentally different way of assessing how memory may influence sentence complexity, one that shifts away from fixed capacity limits to explanations that emphasize the ability to discriminate between which items need to be retrieved from memory during sentence processing. As such, to fully understand the interaction between memory constraints and language, one requires not only well-defined linguistic theory, but also a well-motivated theory of memory operations and architecture.

16.3.1 Candidate memory operations

A variety of cognitive mechanisms and architectures could theoretically play a role in language comprehension (see also Foraker and McElree 2011). Evidence from speed-accuracy tradeoff modeling has supported a theory of memory access during sentence processing that involves *direct-access retrieval*. In this model, memory retrieval involves matching a set of retrieval cues against items in memory. The cues available at retrieval enable access to content-addressable memory representations in one step (hence, direct). Content-addressability means that cues at the retrieval site make contact with memory representations that have overlapping content (McElree and Dosher 1989, 1993; McElree 1996; 1998, 2006; Öztekin and McElree 2007, 2010), and direct-access means retrieval can proceed without recourse to search through extraneous or unrelated memories for the to-be-retrieved item (e.g., Clark and Gronlund 1996; Kohonen 1984). That is, the cues available at the point of retrieval resonate with items in memory according to the amount of (partially) matching content, and the item retrieved is the one with the most overlap or best fit (e.g., Ratcliff 1978).

Perhaps the most notable advantage of this type of memory mechanism is that it enables the rapid recovery of past representations, without introducing the distance-dependent processing time cost found in search operations needed to recover relational information between items (e.g., McElree and Dosher 1993). However, equally notable is the disadvantage that cue-driven direct-access operations are highly susceptible to interference from other constituents in memory that match the cues used for retrieval. Basic memory research indicates that similarity in memory creates retrieval interference through *cue-overload*, where retrieval cues cannot reliably elicit any single target because they are associated with other items in memory (e.g., Öztekin and McElree 2007, 2010; Nairne 2002a, 2002b; Watkins and Watkins 1975).

Alternative candidate memory architectures include serial search, parallel search, and active maintenance. The key prediction of a search operation is that processing time is a function of the number of items in the memory set that must be searched prior to a response. *Serial search* retrieval is a one-by-one, relatively slow search that is necessary for recovering order information, such as the recency of elements in time or across space, and produces intercept differences (Gronlund, Edwards, and Ohrt 1997; McElree 2001, 2006; McElree and Dosher 1993). In *parallel search* retrieval, possible target items are accessed in memory at the same point in time, but produce functions that differ based on the rate of information accrual for the parallel comparisons (Murdock 1971; Townsend and Ashby 1983).

Another fundamental cognitive operation involved in comprehension is *active maintenance*. Modern conceptions of the memory system include controlled attention, where one's focus of attention is an extremely limited-capacity state into and out of which information is shunted very quickly. Several lines of evidence derived from a variety of cognitive and perceptual tasks indicate that a very limited amount of information can be maintained in focal attention (3-4 units: Cowan 2001, 2005; 1 unit: McElree 1998, 2001, 2006). McElree's (2006) conception states that focal attention is just one processing chunk which is quickly replaced by the next chunk of information in mental processing (McElree, 1998; Öztekin, Davachi, and McElree 2010). What constitutes a memory chunk in sentence comprehension is currently underspecified, as it could denote a word, phrase, clause, or potentially (though unlikely) larger stretches of text. In their computational implementation of cue-based retrieval, Lewis and Vasishth (2005) assumed maximal projections constituted a single chunk in memory.

16.3.2 The nature of content-addressable cues

Research within the cue-based framework of language comprehension (Jäger et al. 2017; Lewis et al. 2006; Martin 2016; McElree 2000; McElree et al. 2003; Nicenboim and Vasishth 2018) suggests that representations formed during sentence processing are content-addressable. That is, memory retrieval during language comprehension involves matching a set of retrieval cues against items in memory. The item that provides the best match is then retrieved. In sentence processing, retrieval cues can be generated by (at least) phonological, morphosyntactic, lexical, syntactic, semantic, pragmatic, or discourse information.

As noted above, however, accessing memory in this way leads to the possibility of similarity-based retrieval interference, when multiple items partially match the cues available at the retrieval site. In such cases, discrimination between an intended retrieval target and competitors becomes more difficult. To illustrate this principle, consider sentences (1)-(4), once again. In (1), the subject and verb are adjacent to one another, providing optimal conditions for incrementally building a representation of the complement clause for the matrix verb ‘realize’: having just processed the subject NP ‘the boy’, the comprehender can immediately match it with the final verb ‘yelled’. In contrast, (2) is more challenging to process since ‘the doctor realized’ is now a relative clause intervening between ‘the boy’ and ‘yelled’. Processing the interpolated material will displace the subject NP from active processing, necessitating a retrieval operation to restore ‘the boy’ to active processing when ‘yelled’ is encountered (McElree et al. 2003).

Here, ‘the boy’ and ‘the doctor’ share syntactic and semantic features making retrieval more difficult (syntactic subject of its clause, agent role, and animacy), as well as other aspects of representational similarity (number and gender, stereotypically male for ‘doctor’), which may impinge on forming the correct dependency and final interpretation. In (3) and (4), an additional relative clause is inserted (‘who calmed the mother’ or ‘who ordered a blood test’), but notice that in (3), overlap of the animacy feature between ‘the mother’ and ‘the boy’ entails another source of interference, while the inanimate ‘blood test’ in (4) does not. In these ways, similarity-based interference at retrieval is a necessary by-product of the way linguistic memory is hypothesized to be accessed. To reiterate, memory operations affect the availability and quality of information needed for language comprehension. Whether interpreting spoken, written, or signed language, comprehenders must reconstruct linguistic relationships among the sequentially presented elements that encode meaning.

16.3.3 SAT predictions for memory operations

Recall that asymptote differences in SAT fits reflect the likelihood that an acceptable interpretation is computed or the degree of acceptability of the interpretation. Many factors can contribute to acceptability. A higher asymptote could be due to the higher likelihood of successfully retrieving a representation of a constituent that is sufficient to resolve a non-adjacent dependency (e.g., a subject for a verb, a filler for a gap, an antecedent for an ellipsis or pronoun), or due to the interpretation of the stimulus in a given condition being more plausible or natural than in another condition. Lower asymptotes can be interpreted as a reduction in the quality of retrieved information or as a failure to retrieve the required constituent on a proportion of trials. This includes failed retrieval attempts, including cases where retrieval of an inappropriate item leads to an anomalous interpretation. On some trials, an incorrect first retrieval attempt that results in an inappropriate representation or a problematic one can be followed up with another retrieval attempt that produces an acceptable interpretation – in this way, averaging over several trials contributes to an overall lower asymptote for that condition.

One must keep in mind that inferring causes of asymptote differences is constrained asymmetrically: if the likelihood of successful retrieval is low, then empirical *d′* and estimates of asymptote accuracy will also be low, but if *d′* and asymptote are low, it does not mean that the decrease comes from the retrieval process alone. Interpretation and retrieval cannot be orthogonally dissociated through SAT modelling; only through the careful design of stimuli that differ only in variables hypothesized to affect retrieval but not subsequent interpretation can inferences purely about retrieval be drawn from differences in *d′* or asymptote. However, inferences about retrieval can be made should there be no difference in the speed of processing between conditions, because that is indicative of the content-addressable direct-access retrieval mechanism. Hence, rates and intercepts should not differ despite variation in asymptotic accuracy if the direct-access mechanism is at work.

For serial search, the one-by-one, iterative process produces a linear function, and in SAT, the crucial prediction is that intercepts should increase in time as a function of the number of items that must be searched prior to finding a match (e.g., McElree and Dosher 1989, 1993; Neath 1993; Neath and Knoedler 1994; Öztekin, McElree, Staresina, and Davachi 2008; Sternberg 1975). Parallel search, on the other hand, predicts that only the SAT rates would reflect a speed of processing difference. Parallel search is distinguishable from direct-access retrieval because it predicts no difference in finishing times for positive decisions (*d′* is based on hits and false alarms, which are both ‘yes/acceptable’ decisions) as a function of set size, resulting in a linear relationship between reaction time and set size (Murdock 1971). For language processing, then, SAT rates should decrease systematically as the number of potential binding comparisons at stake increases.

Inasmuch as the hierarchical structure of a sentence is often encoded by the order of constituents within a string, predominantly so in languages such as English, one could argue that a serial search like that used to retrieve recency information might be required to access the elements involved in non-adjacent dependencies. Several SAT experiments have placed interpolated material between the to-be-retrieved constituent and the site of the dependency to test for a backward search mechanism (Martin and McElree 2008, 2009, 2011; McElree 2000; McElree et al. 2003; Van Dyke and McElree 2011). A few experiments have also added additional material before the to-be-retrieved constituent to test for a forward search, in which the search starts at the beginning of an expression (Martin and McElree 2009, 2011; Van Dyke and McElree 2006, 2011).

McElree (1998, 2001, 2006), following Wickelgren, Corbett, and Dosher (1980), argued that measures of the speed of accessing information provide the most direct and unequivocal evidence for whether an item is represented in an active state versus passive state in memory. Measures of processing speed in several cognitive tasks have shown a sharply dichotomous pattern for information in focal attention versus memory, with processing speed being exceptionally fast for responses based on information actively maintained in awareness (Dosher 1981; McElree 1996, 1998, 2001, 2006; McElree and Dosher 1989, 1993; McElree et al. 2003; Öztekin and McElree 2007; Wickelgren et al. 1980). In sentence processing, one would expect to see, for example, notably faster processing on the final verb in (1) above as compared to (2)-(4) if the subject NP were actively maintained when encountering the verb. A comparatively slower speed of accessing the subject NP at the final verb for (2)-(4), and that speed being the same for all three conditions, would indicate a direct-access retrieval operation. Decreasing speed as dependency length increases across (2)-(4) would on the other hand be indicative of a serial search.

16.4 SAT evidence for interactions between syntax and memory operations

McElree (2000) examined structures where increasing surface distance between dependent elements would predict increasing serial search time. In (5), the direct object noun phrase (‘the book’) of a final verb (‘admired/*amused’) was fronted to the beginning of the sentence in a cleft construction. The acceptability of the direct object as a theme of this final verb was manipulated, to be either acceptable (‘the book that the editor admired’) or unacceptable (‘the book that the editor *amused’). The distance between the NP and verb was increased by adding one (6) or two (7) subject-relative clauses. The retrieval site at which the dependency needs to be resolved is the final verb. Participants in the experiment judged sentences using a single-response SAT procedure in which they made an acceptability judgment response at one of six different response times, from 50ms to 3000ms following presentation of the sentence final verb.

(5) This was the book that the editor admired (*amused).

(6) This was the book that the editor who the receptionist married admired (*amused).

(7) This was the book that the editor who the receptionist who quit married admired (*amused).

The best fit SAT function indicated that the asymptotes decreased progressively with more interpolated material, indicating a progressively lower probability of computing a correct interpretation of a sentence, consistent with a decreasing likelihood of retrieving the correct argument from memory. However, the speed of comprehension (rate and intercept) was unaffected by the amount of material intervening between the dependent elements, arguing against backward serial search.

In a next step, McElree et al. (2003) tested whether hierarchical distance rather than surface distance determines search time. Embedded complement clauses differ from center-embedded subject relative clauses in that they increase not only the surface distance between the verb and its argument but also the distance along the right edge of a hierarchical structure (see McElree et al. 2003). In (8), the object NP ('the scandal') is clefted out of its canonical position adjacent to the verb, while in (9) and (10), additional complement clauses are embedded between the clefted NP and the final verb.

(8) It was the scandal that the celebrity relished (*panicked).

(9) It was the scandal that the model believed that the celebrity relished (*panicked).

(10) It was the scandal that the model believed that the journalist reported that the celebrity relished (*panicked).

Once again, accuracy declined progressively as the distance between the dependent elements increased, while the speed of processing remained constant.

In a second experiment, subject-verb dependencies were examined, contrasting cases where the elements were adjacent to one another ('The book __ ripped/*laughed') with cases of intervening material of varying syntactic and semantic overlap: an object relative clause ('that the editor admired'), prepositional phrase plus object relative clause ('from the prestigious press that the editor admired'), an object relative plus subject relative clause ('that the editor who quit the journal admired'), or two object-relative clauses ('that the editor who the receptionist married admired') intervened. When the verb was adjacent to its subject, processing speed was exceptionally fast, consistent with basic memory studies (McElree 2006) indicating that the last item processed was still active in focal attention. The increasing amount and complexity of interpolated material decreased accuracy systematically, consistent with a cue-combination, direct-access operation. Again, the speed of processing did not systematically slow with the amount of interpolated material, counter a serial search operation.

Another way to test for serial search performed over syntactic structure in an iterative manner is through sluicing structures. Martin and McElree (2011) manipulated the number of syntactically available antecedents (11 & 12 one, 13 & 14 two), as well as the distance between the antecedent, 'studied', and sluice site, 'what' (11 & 13 recent, 12 & 14 distant).

(11) In the morning, Michael studied but he didn't tell me what.

(12) Michael studied in the morning, but he didn't tell me what.

(13) Michael slept and studied, but he didn't tell me what.

(14) Michael studied and slept, but he didn't tell me what.

In (13) and (14), both the correct antecedent 'studied' and the incorrect verb 'slept' are syntactically licensed. If search is syntactically constrained, the presence of 'slept' should slow the speed of interpretation compared to (11) and (12). As well, if syntactically-guided search occurs in a forward fashion, then (13) should be slower than (14), and vice versa if it is a backward search. The results, however, revealed no difference in the speed of processing, contra syntactically-guided search, either forward or backward. Instead, the asymptotic differences supported straightforward direct-access retrieval.

Note that these results do not support the conclusion that syntactic structure is not important during retrieval and interpretation of long-distance dependencies. Rather, that the pattern of results is consistent with the engagement of a direct-access retrieval mechanism simply means that it is unlikely that syntactic structures are being serially scanned in order to access antecedent or extracted or dislocated constituents. Similarly, Martin and McElree (2008) found that the number of words and phrases between an antecedent and its ellipsis decreased the likelihood of successful retrieval and interpretation, but did not affect time course to retrieve and interpret the antecedent at the ellipsis site. Secondly, increasing the length and complexity of the antecedent had a similar effect such that only asymptotic accuracy was affected. Again, this pattern of results suggests that syntactic relations between antecedent and ellipsis, while clearly important, do not need to be serially scanned or iteratively evaluated during long-distance dependency resolution. These results by no means suggest that only semantic features are at play during retrieval, nor that syntax is not used during retrieval (see discussion of Van Dyke and McElree 2011, below for evidence of syntactic cues). Martin and McElree (2008) account for these results by positing a pointer mechanism that can point to extant structures in memory without iteratively evaluating them or recomputing them.

Syntactic complexity can alternatively be increased by the number of constituents being bound and interpreted at a retrieval point. Direct-access cues may not be sufficient when interpretation explicitly depends on the relative ordering of constituents. Memory research indicates that a (serial) search is required when relational information is at issue (McElree, 2006). McElree et al. (2003) examined the dependency between a direct object noun (‘the album') and a verb particle ('spread open'), with short (15) and long (16) distances between the constituents. Cases such as (17) and (18) examined variants in which the processing of a verb particle with two arguments ('mount in') required resolving two non-adjacent dependencies ('the album' and 'the stamps') to construct the ditransitive verb phrase ('…mount the stamps in the album'). The unacceptable versions reversed the order of the arguments, resulting in an anomalous interpretation (e.g., '…mount the album in the stamps').

(15) This is the album that the customer found difficult to spread open.

(16) This is the album that the customer who obviously angered the fussy collector found difficult to spread open.

(17) This is the album that the stamps were difficult to mount in.

(18) This is the album that the stamps which obviously angered the fussy collector were difficult to mount in.

Distance served to lower asymptotic accuracy only, for both single- and double-argument sentences. However, single-argument sentences were processed faster than double-argument sentences (earlier intercept and faster rate), demonstrating that resolving two arguments at the one retrieval site required additional time. One explanation of this effect is that relational order information is needed to resolve a dependency when more than one constituent is being bound and interpreted at the retrieval site.

Syntactic role information can also act as a constraining cue at the retrieval site, and appears to have priority over semantic and pragmatic properties. Van Dyke and McElree (2011) compared the interpretation of sentences with differing syntactic contexts, in addition to semantic cues (in/animacy restrictions of the verb). In one experiment, the interfering material matched the syntactic cues at the verb, appearing as a syntactic subject ('motion' or 'witness' in 19), while in a second experiment, the interfering material did not match, appearing in syntactic object position (20).

(19) The attorney who the judge realized had declared that the motion/witness was inappropriate compromised.

(20) The attorney who the judge realized had rejected the motion/witness in the case compromised.

They found retroactive interference effects on asymptotes from a semantic competitor, consistent with much other research (McElree 2000; McElree et al. 2003; Martin and McElree 2008, 2009), but only when the interpolated competitor and to-be-retrieved target were both syntactic subjects (19). This provides evidence that the syntactic role of a constituent affects retrieval in comprehension, and that syntactic constraints appear to be weighted more heavily than semantic constraints. Additionally, syntactic constraints may limit potential sources of interference from memory constituents that have semantic properties in common with the target constituent. Hence, constraints from syntax can help counteract similarity-based interference, which is a critical weakness of a content-addressable memory system.

There is also evidence that morphosyntactic information creates retrieval interference, and thus, by inference, is implicated as a retrieval cue during dependency resolution. For example, using electrophysiology, Martin et al. (2012, 2014) found that grammatical gender agreement between noun phrase ellipsis and its antecedent in Spanish is subject to interference when a noun bearing gender morphology occurs within the ellipsis dependency. Although not SAT evidence, these results highlight the importance of morphosyntax as a retrieval cue during long-distance dependency resolution. More broadly, we would like to note that, to our knowledge, no theory of cue-based retrieval minimizes or discounts the role of syntactic structure in the retrieval and interpretation of non-adjacent dependencies. The fact that interference effects have been found with origins from information inside a relative clause merely implies that the language processing architecture can access information in a syntactic configuration that may not be licensed in other situations. This makes sense if forming long-distance dependencies is not an identical process to computing or generating syntactic structure locally.

16.5 Relations to other aspects of theoretical linguistics

The SAT paradigm provides insight into other kinds of questions in the theoretical linguistics literature, including types of processing and issues of representation. Against the background of distinguishing between the quality or probability of accurate interpretation on the one hand, and the time course of processing on the other, SAT experiments can address a range of linguistic concepts.

Informing a broad set of sentence processing models, McElree and Griffith (1995) found that across experiments and types of model fits, thematic role violations ('Some senators offend elections') produced a later intercept or slower rate than violations of either syntactic category ('Some senators repeatedly elections') or subcategorization ('Some senators roar elections'). These data support models of sentence comprehension where both constituent structure and subcategorization components of syntactic representations are accessible before thematic representations, thus constraining serial, cascade, and parallel models. In this experiment, asymptotic differences did not emerge, indicating that the three kinds of constructions were approximately equivalent in the quality and availability of information required to detect each kind of violation.

Another investigation of McElree and Griffith (1998) focused on the time course of filler-gap processing. They contrasted constructions with subcategorization and thematic role violations, such as ‘It was the evidence that the judge assumed the attorney had loathed/*gone/*astonished’, with island violations, such as ‘It was the evidence that the judge rebuked the attorney who loathed’ or ‘It was the attorney who the judge researched the evidence which astonished’. Again, subcategorization violations produced an earlier intercept than thematic role information. Crucially, island violations consistently showed earlier intercepts than the other sources of information, providing clear evidence that global syntactic configuration information guides a parse very early in processing. Models in which the parser is blocked from predicting gap sites within an island are supported (Stowe 1986), while strong first-resort models in which island constraints are treated as a filter applied after a gap is projected, such as the active-filler strategy (Clifton and Frazier 1989), are not.

McElree (1993) examined how the relative frequencies of a verb’s syntactic frames impacts parsing. Overall results indicated that when the syntactic preference of the verb matched the sentence structure (e.g., ‘watched’ in a transitive frame), a higher asymptote emerged, compared to a mismatch (e.g., ‘rushed’ in a transitive frame), but did not affect time course dynamics. Hence, preferred verb frame frequencies exerted an influence due to stronger representations in the mental lexicon, and were applied at similar speeds over the incremental parse. As well, the asymptote differences provided evidence against frame frequency information being applied serially, where a more frequent frame could temporarily suppress a less frequent structure.

Two nuanced time course differences in McElree (1993) are also of interest. First, a slower rate arose in NP-gap strings with an intransitive-preferring verb in a transitive construction. Second, a slower rate occurred for a syntactic garden-path construction. In both of these cases, slower rates are consistent with reanalysis, following on lower asymptotes due to impoverished retrieval cues.

Additional SAT investigations have examined reanalysis and recovery processes in garden-path ambiguous sentences more specifically. Martin and McElree (2018) tested temporarily ambiguous sentences, like ‘The actress sent the jewelry sparkled/arrived/frowned’, with an initial (incorrect) matrix verb interpretation of ‘sent’, versus the correct reduced relative clause interpretation. The verb ‘sparkled’ is a weak cue for the dependency with ‘actress’, as it is more strongly related to the local noun ‘jewelry’, based on higher latent semantic analysis values, while the verb ‘arrived’ is neutral, equally related to each noun, and the verb ‘frowned’ is more strongly related to the subject than the local noun. The 3 x 2 design included unambiguous relative clause conditions: ‘The actress who was sent the jewelry sparkled/arrived/frowned’. Results demonstrated that retrieval cue strength increased interpretation probability (asymptotes) for all sentences, but did not affect the time course. Ambiguity on the other hand, uniformly slowed rates compared to unambiguous sentences. The rate difference is consistent with reanalysis based on additional attempts to retrieve and interpret a subject. Overall, this profile supports accounts that posit representational differences, such as competing lexical or structural representations. It also indicates that ambiguous structures take more time to process, which is due to multiple parsing attempts. Note that the lack of intercept differences argues against separate, additional repair or reanalysis mechanisms at work.

Contrastingly, Bornkessel et al. (2004) presented evidence from an SAT experiment in German indicating that case information and phrase structure, which can be pulled apart in German, interacted during reanalysis. Participants judged sentences that were temporarily ambiguous, where a garden-path analysis was nominative-initial, but a correct interpretation was dative-initial, and compared them to sentences with a correct nominative-initial interpretation. A later intercept was found for the dative-initial sentences compared to the nominative-initial sentences, supporting a reanalysis operation for syntactic structure. Additionally, within the dative-initial conditions, asymptotic accuracy was higher for an object-experiencer verb than a dative active verb, indicating that the case information associated with the object-experiencer verb provided a stronger cue to guide reanalysis.

While interpretation of the time course results from McElree (1993) and Martin and McElree (2018) may seem contradictory with those of Bornkessel et al. (2004), note that the first two showed rate differences, while the latter showed an intercept difference. Recall that rate differences are more compatible with reapplication of a mechanism already at work, such as additional attempts at retrieving needed elements (Martin and McElree 2018; McElree 1993; McElree et al. 2003), more than one gap to be filled (McElree et al. 2003), or building additional semantic structure (McElree et al. 2006). Intercept differences, on the other hand, can provide evidence of an additional operation or separate mechanism at work over processing time, as in a serial or cascaded parsing routine in which one kind of information is computed before another, or in a parallel architecture in which one kind of information takes more time to compute (Bornkessel et al. 2004; Bott et al. 2012; McElree and Griffith, 1995, 1998). Further examination of different kinds of garden-path ambiguities with SAT methods will help to further elucidate sentence processing mechanisms.

Anaphora is another area where SAT methods can be applied to translate formal linguistic claims to cognitive mechanisms that make time course predictions. Foraker and McElree (2007) assessed two accounts of antecedent prominence, comparing a continuum of activation strength to a special cognitive state akin to focal attention. Approaches such as the Focus Memory Framework (Garrod, Freudenthal, and Boyle 1994; Stewart, Pickering, and Sanford 2000) propose that antecedent representations vary along a continuum of activation strength, which is consistent with a higher probability of retrieving a more prominent antecedent, supporting higher accuracy of coreference resolution – but, no time course distinctions. Alternatively, approaches such as Gundel (1999; Gundel, Hedberg, and Zacharski 1993), and to some extent, Centering Theory (Grosz and Sidner 1986; Grosz, Joshi, and Weinstein 1995), posit that discourse factors which increase antecedent prominence place the most salient item in the psychological focus of attention. This claim predicts a faster speed of processing for coreference involving a prominent antecedent.

Foraker and McElree (2007) compared prominent referents (21, 22) to non-prominent ones (23, 24), as well as a pronoun adjacent to its referent (22, 24) vs. distant from its referent (21, 23).

(21) It was the skillful carpenter who repaired the antique dresser. He hammered (*creaked).

(22) What the skillful carpenter repaired was the antique dresser. It creaked (*hammered).

(23) What the skillful carpenter repaired was the antique dresser. He hammered (*creaked).

(24) It was the skillful carpenter who repaired the antique dresser. It creaked (*hammered).

The syntactic clefting structure in (21, 24) makes the noun phrase ‘skillful carpenter’ more prominent, while the pseudo-cleft in (22, 23) renders ‘antique dresser’ more prominent. Unacceptable versions were constructed by switching the last verb, creating an animacy violation during binding with the pronoun.

When the pronoun referred back to a prominent referent, asymptotic accuracy was higher than the non-prominent conditions, consistent with facilitated retrieval of the referent representation. However, prominence did not affect the speed of processing, arguing against active maintenance in a specialized state. Instead, speed of processing was faster when the antecedent and pronoun were adjacent (22, 24), compared to not (21, 23). The faster speed supports an active maintenance explanation for adjacent elements only (see also McElree et al. 2003), not for prominent antecedent conditions.

Structural locality influences on reflexive anaphor resolution have also been examined with SAT modeling (Dillon, Chow, Wagers, Guo, Liu, and Phillips 2014). For the Mandarin Chinese reflexive ‘ziji’, Dillon et al. found an earlier intercept for retrieval and binding of ‘ziji’ with an antecedent within a local syntactic domain, compared to long-distance binding. These results appear to show that retrieval is limited to the local subject position at first (even when that is the dispreferred interpretation overall), suggesting that a subset of features germane to the dependency is used as retrieval cues, rather than all features. Although these results are of interest, we advise additional investigation, as the differences in observed d' were only marginally different yet fit with different asymptotes, and accuracy was extremely low overall. Perhaps other factors may be at play in this case of anaphor processing.

Finally, additional SAT investigations have tested models and theories involving metonymy, metaphor, enriched composition, and scalar implicatures. Metonymic expressions were less likely to be computed than literal controls, with no differences in time course (Bott et al. 2016), supporting direct access to metonymic senses, and arguing against an indirect, literal-first type of model. Similarly, figurative interpretations were less likely to be recovered or computed than literal ones, with no differences in time course (McElree and Nordlie 1999), also arguing against a serial, literal-first approach. Relatedly, enriched composition expressions were less likely to be sensibly computed than non-coerced controls (McElree et al. 2006). However, in this case, enriched expressions also displayed slower rates, consistent with the claim that building additional semantic structure required more time. Finally, pragmatic upper-bound interpretations of *some* scalar implicatures were less likely to be computed than logical, lower-bound interpretations (Bott et al. 2012). Importantly, pragmatic interpretations, which required the scalar implicature, were also consistently slower in time course intercept and rate than logical ones. Bott et al. (2012) discuss several causes of costly implicatures, including extra computations, underinformativeness, and aspects of an inferential mechanism itself, such as implementing the epistemic step.

16.6 Future applications of SAT

Looking to future SAT applications, a number of linguistic theories have posited that anaphora resolution can be resolved in different ways (e.g. Bosch 1983; Grodzinsky and Reinhart 1993; Reuland 2001, 2011). Although the precise characterization of these theories differ, many assume that anaphora resolution can be resolved via either a syntactic route, typically referred to as variable binding, or via discourse-mediated coreference assignment (for discussion, see Reuland 2011). Reuland argued that in cases where both routes are available, an economy principle dictates that variable binding should be computed before coreference assignment. Results from eye-movement studies have provided mixed support for this claim, with some researchers claiming evidence in favour of a preference for variable binding (Koornneef 2008), and others not (Cunnings, Patterson, and Felser 2014). The SAT procedure can provide an explicit way of modelling the likelihood and time course of pronoun resolution to help tease apart the hypothesized dissociation between variable binding and coreference resolution.

Another important issue to discuss in relation to how linguistic representations are accessed from memory relates to c-command. C-command describes the relationship between two constituents in a syntactic tree structure in terms of hierarchical dominance. The standard definition of c-command is that a constituent c-commands its sister constituents, and any constituents that they dominate (Reinhart 1983). C-command is crucial to the linguistic characterization of syntactic constraints on linguistic dependencies. For example, in (25), the traditional linguistic characterization of how 'himself' can be interpreted is that it must be bound by a c-commanding antecedent in the same local domain (Chomsky, 1981).

(25) The boy who Kevin spoke to yesterday morning injured himself.

Cue-based content-addressable memory access relies on features to access information in memory. However, c-command is an inherently relational concept between sentence constituents that cannot be reduced to a feature. We cannot, for example, say that *Kevin* lacks a [+C-COMMAND] feature to restrict it from being retrieved upon encountering the reflexive, because while *Kevin* does not c-command the reflexive, it does c-command other constituents in the sentence ('spoke to yesterday morning'). Other linguistic dependencies are also typically described as being restricted by c-command.

Given the relational nature of c-command and its importance in constraining linguistic dependencies, it might be surprising that evidence from the SAT paradigm suggests language comprehension involves memory access via feature-based direct-access retrieval rather than, for example, a serial search that could utilize relational information (McElree et al. 2003). One way to address this issue has been to devise feature-based proxies that encode the c-command relation via a set of features on items in memory that can subsequently be utilised during cue-based retrieval (e.g. Cunnings et al. 2014; Kush 2013; Kush, Lidz, and Phillips 2015). However, on a theoretical level, it is important to note that these feature-based proxies are not c-command, as they are inherently non-relational. Existing research on the role of c-command in constraining memory access during processing has typically assumed that while feature-based proxies might be utilized as retrieval cues to guide memory access online, the linguistic characterization of constraints on linguistic dependencies can still be described in terms of a c-command relation. However, a more radical and controversial conclusion might be that, if it is indeed the case that memory access during language comprehension relies on cue-based retrieval, and on the assumption of a tight relationship between the grammar and the parser, constraints on linguistic dependencies should rather be theoretically characterized in terms of content-based features, instead of relational notions such as c-command. Other possibilities are that the relational order aspect of c-command is not accessed during the retrieval and interpretation of non-adjacent dependencies, or is a much weaker or less reliable constraint that carries lower weight, or that interacts with syntactic role, such as subjecthood (Van Dyke and McElree 2011). A fundamentally different way of conceiving of the relationship between grammar (or grammatical constraints like c-command) and the parser (or computation and its behavior consequence during processing) is to see grammar as a system of (neural) state spaces that the network can enter into (Martin, 2016). On this view, grammar is implicitly represented (see Rust 2014 for a discussion of how implicit information can become explicit and accessible in neural systems) and no additional parsing mechanism is needed. It is only representation of the information that determines which state the system enters next, such that detection of representational state through sensory signals essentially replaces the role of the parser (Martin 2016).

As already noted, evidence from the SAT paradigm indicates that language comprehension is susceptible to retrieval interference, and that this interference is dependent on the similarity between items in memory. Similarity-based retrieval interference is in ways similar to the concept of relativized minimality in the theoretical linguistics literature (Rizzi 1990, 2011). Relativized minimality states that a linguistic dependency between a displaced constituent and its canonical sentence position can be disrupted when a c-commanding constituent, whose morphosyntactic features match that of the displaced constituent, intervenes. Although the notions of similarity-based interference and relativized minimality have been formalized to account for different linguistic phenomena, both predict that the success of linguistic dependency resolution is influenced by the similarity between sentence constituents. Cue-based retrieval provides a processing implementation for relativized minimality that is rooted in the principles of human recognition memory, and offers a mechanistic explanation for why intervention effects occur: they fall out naturally as a consequence of the way memory is accessed in cue-based retrieval. Precise characterization of constraints on how memory is accessed during linguistic dependency resolution may help provide a unifying bridge between work in theoretical linguistics on the characterization of linguistic constraints on dependency resolution, and work in psycholinguistics on the time-course of memory access during sentence processing. Future research using the SAT paradigm will help formalize this link between memory access and linguistic representation.

16.7 Conclusion

While many different experimental paradigms are available to the linguist and psycholinguist interested in investigating linguistic representation and processing, the SAT procedure provides the best means to veridically estimate the tradeoff between speed and accuracy during language processing, which provides a comprehensive picture of when and how interpretation develops over processing time. In this chapter, we explained how the SAT paradigm can provide clear evidence about time course of processing that is unconfounded by accuracy or probability of interpretation. We also described the prominent role that SAT evidence has taken in integrating memory models into psycholinguistic theory, and reviewed how SAT evidence can be used to inform other issues of debate in linguistics and psycholinguistics, which we hope will inspire future use of the SAT paradigm within the experimental syntax literature.

Bibliography

Bornkessel, Ina, Brian McElree, Matthias Schlesewsky, and Angela D. Friederici (2004). 'Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking', *Journal of Memory and Language* 51.4: 495-522.

Bösch, Peter (1983). *Agreement and anaphora: A study of the role of pronouns in syntax and discourse*. London: Academic Press.

Bott, Lewis, Todd M. Bailey, and Daniel Grodner (2012). 'Distinguishing speed from accuracy in scalar implicatures', *Journal of Memory and Language* 66: 123-142.

Bott, Lewis, Alice Rees, and Steven Frisson (2016). 'The time course of familiar metonymy', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42: 1160-1170.

Chandler, John P. (1969). 'Subroutine STEPIT – finds local minimum of a smooth function of several parameters', *Behavioral Science* 14: 81-82.

Clark, Steven E., and Scott D. Gronlund (1996). 'Global matching models of recognition memory: How the models match the data', *Psychonomic Bulletin & Review* 3: 37-60.

Clifton, Charles, and Lyn Frazier (1989). 'Comprehending sentences with long-distance dependencies', *Linguistic structure in language processing*. Springer, Dordrecht. 273-317.

Chomsky, Noam (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Cohen, Neal J., and Howard Eichenbaum (1993). *Memory, amnesia, and the hippocampal system.* Cambridge, MA: MIT Press.

Cowan, Nelson (2001). 'The magical number 4 in short-term memory: A reconsideration of mental storage capacity', *Behavioral and Brain Sciences* 24: 87-185.

Cowan, Nelson (2005). *Working memory capacity (Essays in cognitive psychology).* New York: Psychology Press.

Cunnings, Ian, Clare Patterson, and Claudia Felser (2014). ‘Variable binding and coreference in sentence comprehension: evidence from eye movements’, *Journal of Memory and Language* 71.1: 39-56.

Daneman, Meredyth, and Patricia A. Carpenter (1980). ‘Individual differences in working memory and reading’, *Journal of Verbal Learning and Verbal Behavior* 19.4: 450-466.

Dillon, Brian, Wing-Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu, and Colin Phillips (2014). ‘The structure-sensitivity of memory access: evidence from Mandarin Chinese’, *Frontiers in Psychology* 5: Article 1025.

Donders, Franciscus Cornelis (1969). ‘On the speed of mental processes’, *Attention and Performance II, Acta Pscyhologica* 30: 412-431. ed. and trans. by W. G. Koster. Amsterdam: North-Holland. (Original work published in 1868)

Dosher, Barbara A. (1981). ‘The effect of delay and interference: A speed-accuracy study’, *Cognitive Psychology* 13: 551-582.

Foraker, Stephani, and Brian McElree (2007). ‘The role of prominence in pronoun resolution: active versus passive representations’, *Journal of Memory and Language* 56.3: 357-383.

Foraker, Stephani, and Brian McElree (2011). ‘Comprehension of linguistic dependencies: speed-accuracy tradeoff evidence for direct-access retrieval from memory’, *Language and Linguistics Compass* 5.11: 764-783.

Garrod, Simon, Daniel Freudenthal, and Elizabeth Boyle (1994). ‘The role of different types of anaphor in the on-line resolution of sentences in a discourse’, *Journal of Memory and Language* 33: 39-68.

Gibson, Edward (2000). ‘The dependency locality theory: A distance-based theory of linguistic complexity’, in *Image, Language, Brain*. 95-126.

Gordon, Peter C., Randall Hendrick, and Marcus Johnson (2001). ‘Memory interference during language processing’, *Journal of Experimental Psychology: Learning, Memory and Cognition* 27: 1411-1423.

Gordon, Peter C., Randall Hendrick, and Marcus Johnson (2004). ‘Effects of noun phrase type on sentence complexity’, *Journal of Memory and Language* 51: 97-114.

Gordon, Peter C., Randall Hendrick, and William H. Levine (2002). ‘Memory-load interference in syntactic processing’, *Psychological Science* 13: 425-430.

Gordon, Peter C., Randall Hendrick, Marcus Johnson, and Yoonhyoung Lee (2006). ‘Similarity-based interference during language comprehension: evidence from eye tracking during reading’, *Journal of Experimental Psychology: Learning, Memory and Cognition* 32: 1304-1321.

Grodner, Daniel, and Edward Gibson (2005). ‘Consequences of the serial nature of linguistic input for sentential complexity’, *Cognitive Science* 29.2: 261-290.

Grodzinsky, Yosef, and Tanya Reinhart (1993). ‘The innateness of binding and coreference’, *Linguistic Inquiry* 24.1: 69-101.

Gronlund, Scott D., Mark B. Edwards, and Daryl D. Ohrt (1997). ‘Comparison of the retrieval of item versus spatial position information’, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23: 1261-1274.

Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein (1995). ‘Centering: a framework for modelling the local coherence of discourse’, *Computational Linguistics* 21.2: 203–226.

Grosz, Barbara J., and Candace L. Sidner (1986). ‘Attention, intentions, and the structure of discourse’, *Computational Linguistics* 12.3: 175–204.

Gundel, Jeanette K. (1999). ‘On different kinds of focus’, in P. Bosch and R. van der Sandt, (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives.* Cambridge: Cambridge University Press, 293-305.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski (1993). ‘Cognitive status and the form of referring expressions in discourse’, *Language* 69: 274-307.

Hagoort, Peter (2003). ‘How the brain solves the binding problem for language: a neurocomputational model of syntactic processing’, *Neuroimage* 20: S18-S29.

Jäger, Lena A., Felix Engelmann, and Shravan Vasishth. (2017). ‘Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis’, *Journal of Memory and Language* 94: 316-339.

James, William (1918). *The Principles of Psychology*. New York, NY: Holt. (Original work published 1890)

Judd, Charles M., and Gary H. McClelland (1989). *Data analysis: A model-comparison approach*. San Diego: Harcourt Brace.

Kohonen, Teuvo (1984). *Self-organization and associative memory.* New York, NY: Springer.

Koornneef, Arnout Willem (2008). *Eye-catching anaphora.* Dissertation. Utrecht University, The Netherlands.

Kush, Dave W. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing*. Dissertation. University of Maryland, College Park, MD.

Kush, Dave, Jeffrey Lidz, and Colin Phillips (2015). ‘Relation-sensitive retrieval: evidence from bound variable pronouns’, *Journal of Memory and Language* 82: 18-40.

Lewis, Richard L., and Shravan Vasishth (2005). ‘An activation-based model of sentence processing as skilled memory retrieval’, *Cognitive Science* 29: 375-419.

Lewis, Richard L., Shravan Vasishth, and Julie A. Van Dyke (2006). ‘Computational principles of working memory in sentence comprehension’, *Trends in Cognitive Science* 10: 445-454.

Martin, Andrea E. (2016). ‘Language processing as cue integration: grounding the psychology of language in perception and neurophysiology’, *Frontiers in Psychology* 7: Article 120.

Martin, Andrea E., and Brian McElree (2008). ‘A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis’, *Journal of Memory and Language* 58: 879-906.

Martin, Andrea E., and Brian McElree (2009). ‘Memory operations that support language comprehension: evidence from verb-phrase ellipsis’, *Journal of Experimental Psychology: Learning Memory & Cognition* 35: 1231-1239.

Martin, Andrea E., and Brian McElree (2011). ‘Direct-access retrieval during sentence comprehension: Evidence from sluicing’, *Journal of Memory and Language* 64: 327-343.

Martin, Andrea E., and Brian McElree (2018). ‘Retrieval cues and syntactic ambiguity resolution: speed-accuracy tradeoff evidence’, *Language, Cognition, and Neuroscience* 33: 769-783.

Martin, Andrea. E., Mante S. Nieuwland, and Manuel Carreiras (2012). ‘Event-related brain potentials index cue-based retrieval interference during sentence comprehension,’ *Neuroimage* 59.2: 1859-1869.

Martin, Andrea E., Mante S. Nieuwland, and Manuel Carreiras, M. (2014). ‘Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis,’ *Brain and Language* 135: 42-51.

McElree, Brian (1993). ‘The locus of lexical preference effects in sentence comprehension: a time-course analysis’, *Journal of Memory and Language* 32: 536-571.

McElree, Brian (1996). ‘Accessing short-term memory with semantic and phonological information: A time-course analysis’, *Memory & Cognition* 24: 173-187.

McElree, Brian (1998). ‘Attended and non-attended states in working memory: accessing categorized structures’, *Journal of Memory and Language* 38: 225-252.

McElree, Brian (2000). ‘Sentence comprehension is mediated by content-addressable memory structures’, *Journal of Psycholinguistic Research* 29: 111-123.

McElree, Brian (2001). ‘Working memory and focal attention’, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27: 817-835.

McElree, Brian (2006). ‘Accessing recent events’, in B. H. Ross (ed.), *The psychology of learning and motivation Vol. 46*. San Diego: Academic Press.

McElree, Brian, and Barbara A. Dosher (1989). ‘Serial position and set size in short term memory: the time course of recognition’, *Journal of Experimental Psychology: General* 118: 346-373.

McElree, Brian, and Barbara A. Dosher (1993). ‘Serial retrieval processes in the recovery of order information’, *Journal of Experimental Psychology: General* 122: 291-315.

McElree, Brian, Stephani Foraker, and Lisbeth Dyer (2003). ‘Memory structures that subserve sentence comprehension’, *Journal of Memory and Language* 48: 67-91.

McElree, Brian, and Teresa Griffith (1995). ‘Syntactic and thematic processing in sentence comprehension: evidence for a temporal dissociation’, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21.1: 134-157.

McElree, Brian, and Teresa Griffith (1998). ‘Structural and lexical constraints on filling gaps during sentence comprehension: a time-course analysis’, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.2: 432-460.

McElree, Brian, and Johanna Nordlie (1999). ‘Literal and figurative interpretations are computed in equal time’, *Psychonomic Bulletin & Review* 6: 486-494.

McElree, Brian, Liina Pylkkänen, Martin Pickering, and Matthew J. Traxler (2006). ‘A time course analysis of enriched composition’, *Psychonomic Bulletin & Review* 13: 53-59.

Miller, George, and Noam Chomsky (1963). ‘Finitary models of language users’, in D. R. Luce, R. R. Bush, & E. Galanter (eds.), *Handbook of Mathematical Psychology (Vol. 2)*. New York: John Wiley.

Murdock, Bennet B. Jr. (1971). ‘A parallel-processing model for scanning’, *Perception and Psychophysics* 10: 289-291.

Nairne, James S. (2002a). ‘The myth of the encoding-retrieval match’, *Memory* 1: 389-395.

Nairne, James S. (2002b). ‘Remembering over the short-term: the case against the standard model’, *Annual Review of Psychology* 53.1: 53-81.

Neath, Ian. (1993). ‘Distinctiveness and serial position effects in recognition’, *Memory & Cognition* 21: 689-698.

Neath, Ian, and Alicia J. Knoedler (1994). ‘Distinctiveness and serial position effects in recognition and sentence processing’, *Journal of Memory and Language* 33.6: 776-795.

Nicenboim, Bruno, and Shravan Vasishth. (2018). ‘Models of retrieval in sentence comprehension’, *Journal of Memory and Language* 99: 1-34.

Oberauer, Klaus (2002). ‘Access to information in working memory: exploring the focus of attention’, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28: 411-421.

Öztekin, Ilke, and Brian McElree (2007). ‘Proactive interference slows recognition by eliminating fast assessments of familiarity’, *Journal of Memory and Language* 57: 126-149.

Öztekin, Ilke, and Brian McElree (2010). ‘Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution’, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36.2: 383-397.

Öztekin, Ilke, Brian McElree, Bernhard P. Staresina, and Lila Davachi (2008). ‘Working memory retrieval: contributions of the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus’, *Journal of Cognitive Neuroscience* 21: 581-593.

Öztekin, Ilke, Lila Davachi, and Brian McElree (2010). ‘Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store’, *Psychological Science* 21.8: 1123-1133.

Parker, Dan. (2015). ‘Two is not always better than one: modeling evidence for a single structure building system’. Poster at the Architectures & Mechanisms for Language Processing (AMLaP) 2015 Conference. University of Malta.

Parker, Dan, Michael Shvartsman, and Julie A. Van Dyke (2017). ‘The cue-based retrieval theory of sentence comprehension: new findings and new challenges’, in L. Escobar, V. Torrens, and T. Parodi (eds.) *Language Processing and Disorders*. Newcastle: Cambridge Scholars Publishing, 121-144.

Phillips, Colin, Matthew W. Wagers, and Ellen F. Lau (2011). ‘Grammatical illusions and selective fallibility in real-time language comprehension’, *Experiments at the Interfaces* 37: 147-180.

Ratcliff, Roger (1978). ‘A theory of memory retrieval’, *Psychological Review* 85: 59-108.

Reed, Adam V. (1973). ‘Speed–accuracy trade-off in recognition memory’, *Science* 181: 574–576.

Reed, Adam V. (1976). ‘List length and the time course of recognition in immediate memory’, *Memory & Cognition* 4: 16-30.

Reinhart, Tanya. (1993). ‘Coreference and bound anaphora: a restatement of the anaphora questions’, *Linguistics and Philosophy* 6.1: 47-88.

Reuland, Eric. (2001). ‘Primitives of binding’, *Linguistic Inquiry* 32.3: 439-492.

Reuland, Eric J. (2011). *Anaphora and language design*. Cambridge, MA: MIT Press.

Rizzi, Luigi. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi (2011). ‘Minimality’, in C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*. New York/Oxford: Oxford University Press, 220–238.

Rust, Nicole C. (2014). ‘Population-based representations: from implicit to explicit’, in M. S. Gazzaniga, G. R. Mangun, and S.-J. Blakemore (eds.), *The Cognitive Neurosciences*. Cambridge, MA: MIT Press, 337-348.

Sprouse, Jon, Carson T. Schütze, and Diogo Almeida. (2013). ‘A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010’, *Lingua* 134: 219-248.

Sternberg, Saul. (1969). ‘The discovery of processing stages: extensions of Donders’ method’, *Acta Psychologica* 30: 276-315.

Sternberg, Saul. (1975). ‘Memory scanning: new findings and current controversies’, *Quarterly Journal of Experimental Psychology* 27: 1-32.

Stewart, Andrew J., Martin J. Pickering, and Anthony J. Sanford (2000). ‘The time course of the influence of implicit causality information: focusing versus integration accounts’, *Journal of Memory and Language* 42: 423–443.

Stowe, Laurie A. (1986). 'Parsing wh-constructions: evidence for on-line gap location', *Language and Cognitive Processes* 1: 227-245.

Townsend, James T., and F. Gregory Ashby (1983). *The stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

Van Dyke, Julie A (2007). 'Interference effects from grammatically unavailable constituents during sentence processing', *Journal of Experimental Psychology: Learning, Memory and Cognition* 33: 407-430.

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

Van Dyke, Julie A., Clinton L. Johns, and Anuenue Kukona (2014). 'Low working memory capacity is only spuriously related to poor reading comprehension', *Cognition* 131.3: 373-403.

Van Dyke, Julie A., and Richard L. Lewis (2003). 'Distinguishing effects of structure and decay on attachment and repair: a retrieval interference theory of recovery from misanalyzed ambiguities', *Journal of Memory and Language* 49: 285-316.

Van Dyke, Julie A., and Brian McElree (2006). 'Retrieval interference in sentence comprehension', *Journal of Memory and Language* 55: 157-166.

Van Dyke, Julie A., and Brian McElree (2011). 'Cue-dependent interference in comprehension', *Journal of Memory and Language* 65: 247-263.

Vasishth, Shravan, Sven Brüssow, Richard L. Lewis, and Heiner Drenhaus. (2008). 'Processing polarity: how the ungrammatical intrudes on the grammatical', *Cognitive Science* 32.4: 685-712.

Wagers, Matthew W., Ellen F. Lau, and Colin Phillips. (2009). ‘Agreement attraction in comprehension: representations and processes’, *Journal of Memory and Language* 61.2: 206-237.

Warren, Tessa, and Edward Gibson. (2002). ‘The influence of referential processing on sentence complexity’, *Cognition* 85.1: 79-112.

Watkins, Olga C., and Michael J. Watkins (1975). ‘Build-up of proactive inhibition as a cue overload effect’, *Journal of Experimental Psychology: Human Learning and Memory* 104.4: 442-452.

Wickelgren, Wayne A. (1977). ‘Speed-accuracy tradeoff and information processing dynamics’, *Acta Psychologica* 41: 67-85.

Wickelgren, Wayne A., Albert T. Corbett, and Barbara A. Dosher (1980). ‘Priming and retrieval from short-term memory: a speed-accuracy tradeoff analysis’, *Journal of Verbal Learning and Verbal Behavior* 19: 387-404.