of bounded rationality: Information critical to the successful recognition of phonetic categories often occurs downstream in the speech signal (Bard et al. 1988; Grosjean 1985). Effective language understanding thus requires maintaining and integrating graded support for different phonetic categories provided by a sound’s acoustics (its subcategorial information) with information present in the downstream signal. Indeed, more recent work suggests that comprehension deviates from this. For example, evidence in VOT are not immediately forgotten but are still available downstream at the end of a multisyllabic word (McMurray et al. 2009; see Dalhan 2010, for further discussion of right-context effects).

Of particular relevance is a line of work initiated by Connine et al. (1991, Expt. 1). They manipulated VOT in the initial segment of target words (dent/dentent) and embedded these words in utterances with downstream information about the word’s identity (e.g., “The dent/tent in the fender” or “... forest”). They found that listeners can maintain subcategorical phonetic detail and integrate it with downstream information even beyond word boundaries.

Chunk-and-Pass does not predict these results. Recognizing this, C&C allow violations of Now-or-Never, as long as “such online ‘right-context effects’ [are] highly local, because raw perceptual input will be lost if it is not rapidly identified” (sect. 3.1, para. 7).

This substantially weakens the predictive power of their proposal. On the other hand, Connine et al.’s results do seem to support this qualification. They reported that subcategorical phonetic detail (a) was maintained only 3 syllables downstream, but not 6–8, and (b) was maintained only for maximally ambiguous tokens.

Recent work, however, points to methodological issues that call both of these limitations into question (Bicknell et al. 2015). Regarding (a), Connine et al. allowed listeners to respond at any point in the sentence: On 84% of trials in the 6–8 syllable condition, listeners categorized the target word prior to hearing the relevant right-context (e.g., fender or forest). Therefore, these responses could not probe access to subcategorical information. In a replication that avoided this problem, we found that subcategorical detail decays more slowly than Connine et al.’s analysis would suggest: Subcategorical detail was maintained for at least 6–8 syllables (the longest range investigated). Regarding (b), Connine et al.’s analysis was based on proportions, rather than log-odds. Rational integration of downstream information with subcategorical information should lead to additive effects in log-odds space (which, in proportional space, then are largest around the maximally ambiguous tokens. Bicknell et al. 2015). This is indeed what we found. The effect of downstream information on the log-odds of hearing dent (or tent) was constant across the entire VOT range. In short, subcategorical information is maintained longer than previous studies suggested, not immediately discarded by chunking (see also Szostak & Pitt 2013). Moreover, maintenance is not limited to special cases; it is the default (Brown et al. 2014).

Clearly, language processing is subject to cognitive limitations; many—if not most—theories of language processing acknowledge this. In its general form, the Now-or-Never bottleneck thus embodies an idea as old as the cognitive sciences: that observable behavior and the cognitive representations and mechanisms underlying this behavior are primarily driven by a priori (static/ fixed) cognitive limitations. This contrasts with another view: Cognitive and neural systems have evolved efficient solutions to the computational tasks agents face (Anderson 1990). Both views have been productive, providing explanations for perception, motor control, and cognition, including language (and C&C have contributed to both views). A number of proposals have tied together these insights. This includes the idea of bounded rationality, that is, rational use of limited resources given task constraints (Hovest et al. 2009; Neumann et al. 2014; Simon 1982, for language; e.g., Bicknell & Levy 2010; Feldman et al. 2009; Kleinschmidt & Jaeger 2015; Kuperberg & Jaeger 2016; Lewis et al. 2013). Chunk-and-Pass is a step backward because it blurs the connection between these two principled dimensions of theory development. Consequently, it fails to predict systematic maintenance of subcategorical information, whereas bounded rationality predicts this property of language processing and offers an explanation for it.

The Now-or-Never bottleneck makes novel, testable predictions only insofar as it makes strong claims about comprehenders’ (in)ability to maintain lower-level information beyond the “now.” The studies we summarized above are inconsistent with this claim. Similarly inconsistent is evidence from research on reading suggesting that lower-level information survives long enough to influence incremental parsing (Levy 2011; Levy et al. 2009). Moreover, the history of research on categorical perception provides a word of caution: Rather than focusing too much on cognitive limitations, it is essential for researchers to equally consider the computational problems of language processing and how comprehender goals can be effectively achieved.

Linguistic representations and memory architectures: The devil is in the details

doi:10.1017/S0140525X15000746, e68

Dustin Alfonso Chacón, Shota Momma, and Colin Phillips

Department of Linguistics, University of Maryland, College Park, MD 20742. dustin.alfonso@gmail.com shotam2@gmail.com colin@umd.edu

Abstract: Attempts to explain linguistic phenomena as consequences of memory constraints require detailed specification of linguistic representations and memory architectures alike. We discuss examples of support for locality biases in language comprehension and production, and their link to memory constraints. Findings do not generally favor Christiansen & Chater’s (C&C) approach. We discuss connections to debates that stretch back to the nineteenth century.

It is important to understand how language is shaped by cognitive constraints, and limits on memory are natural culprits. In this regard, Christiansen & Chater (C&C) join a tradition in language research that has a long pedigree (Frazier & Fodor 1978; Wundt 1904) and to which we are sympathetic. C&C’s model aims to integrate an impressive range of phenomena, but the authors play fast and loose with the details; they mischaracterize a number of phenomena; and key predictions depend on auxiliary assumptions that are independent of their model. An approach that takes the details of linguistic representations and memory architectures more seriously will ultimately be more fruitful. We illustrate using examples from comprehension and production.

C&C propose that comprehenders can maintain only a few low-level percepts at once and must therefore quickly encode higher-order, abstract representations. They argue that this explains the pervasive bias for shorter dependencies. However, memory representations are more than simple strings of words that quickly vanish. Sentences are encoded as richly articulated, connected presentations that persist in memory; perhaps without explicit encoding of order, and memory access is similarly articulated (Lewis et al. 2006). As evidence of their model, C&C cite agreement attraction in sentences like The key to the cabinets are on the table. These errors are common in production and often go unnoticed in comprehension, and it is tempting to describe them in terms of “proximity concord” (Quirk et al. 1972). But this is inaccurate. Agreement attraction is widespread in cases where the distractor is further from the verb than the true subject, as in The musicians who the reviewer praise so highly will win (Bock & Miller 1991). Attraction is asymmetrical, yielding “illusions of grammaticality” but not “illusions of ungrammaticality” (Wagers et al. 2009), and depends on whether the distractor is syntactically “active” (Franck et al. 2010). These facts are surprising if attraction reflects simple recency, but they can be captured in a model that combines articulated linguistic representations with a context-addressable memory architecture (Dillon et al. 2013;
McElree et al. 2003). Hence, agreement attraction fits C&C’s broadest objective, deriving attraction from memory constraints, but only if suitably detailed commitments are made.

C&C also endorse the appealing view that locality constraints in syntax (‘island effects’: Ross 1967) can be reduced to memory-driven locality biases in the processing of filler-gap dependencies (Kluender & Kutas 1995). Details matter here, too, and they suggest a different conclusion. When linear and structural locality diverge, as in head-final languages such as Japanese, it becomes clear that the bias for shorter filler-gap dependencies in processing is linear, whereas grammatical locality constraints are structural (Aoshima et al. 2004; Chacón et al., submitted; Omaki et al. 2014).

The moral that we draw from these examples is that each reductionist claim about language must be evaluated on its own merits (Phillips 2013).

Turning to production, C&C argue that incrementality and locality biases reflect severe memory constraints, suggesting that we speak “into the void.” This amounts to what is sometimes called radical incrementality (Ferreira & Swets 2002). It implies that sentence production involves word-by-word planning that is tightly synchronized with articulation – for example, planning is just-in-time, to avoid a bias for local dependencies between words. However, this view of production does not reflect memory constraints alone, and it is empirically unwarranted.

Radical incrementality carries a strong representational assumption whose problems were pointed out in the late nineteenth century. The philologist Hermann Paul, an opponent of Wilhelm Wundt, argued that a sentence is essentially an associative sum of clearly segmentable concepts, each of which can trigger articulation in isolation. Radical incrementality requires this assumption, as it presupposes the isolability of each word or phrase in a sentence at all levels of representation. Memory constraints alone do not require this assumption, and so there is a gap in C&C’s argument that memory constraints entail radical incrementality. Indeed, Wundt was already aware of memory limitations, and yet he adopted the contrasting view that sentence planning involves a successive scanning (upperception) of a sentence that is simultaneously present in the background of consciousness during speech (Wundt 1904). The historical debate illustrates that radical incrementality turns on representational assumptions rather than directly following from memory limitations.

Empirically, radical incrementality has had limited success in accounting for production data. Three bodies of data that C&C cite turn out to not support their view. First, the scope of planning at higher levels (e.g., conceptual) can span a clause (Meyer 1996; Smith & Wheelock 1999). Also, recent evidence suggests that linguistic dependencies can modulate the scope of planning (Lee et al. 2013; Momma et al. 2015, in press). Second, since Wundt’s time, availability effects on word order have not led researchers to assume radical incrementality (see Levelt 2012 for an accessible introduction to Wundt’s views). Bock (1987) emphasized that availability effects on order result from the tendency for accessible words to be assigned a higher grammatical function (e.g., subject). In languages where word order and the grammatical functional hierarchy dissociate, availability effects support the grammatical function explanation rather than radical incrementality (Christianson & Ferreira 2005). Third, contrary to C&C’s claim, early observations about speech errors indicated that exchange errors readily cross phrasal and clausal boundaries (Garrett 1980).

C&C could argue that their view is compatible with many of these data; memory capacity at higher levels of representation is left as a free parameter. But this is precisely the limitation of their model: Specific predictions depend on specific commitments. Radical incrementality is certainly possible in some circumstances, but it is not required, and this is unexpected under C&C’s view that speaking reduces to a chain of word productions that are constrained by severe memory limitations.

To conclude, we affirm the need to closely link language processes and cognitive constraints, and we suspect the rest of the field does too. However, the specifics of the memory system and linguistic representations are essential for an empirically informative theory, and they are often validated by the counterintuitive facts that they explain.

**Gestalt-like representations hijack Chunk-and-Pass processing**

doi:10.1017/S0140525X15000758, e69

Magda L. Dumitru
Department of Cognitive Science, Macquarie University, Sydney NSW 2109, Australia.

magda.dumitru@gmail.com

**Abstract:** Christiansen & Chater (C&C) make two related and somewhat contradictory claims, namely that the ever abstract language representations built during Chunk-and-Pass processing allow for ever greater interference from extra-linguistic information, and that it is nevertheless the language system that re-codes incoming information into abstract representations. I analyse these claims and discuss evidence suggesting that Gestalt-like representations hijack Chunk-and-Pass processing.

Christiansen & Chater (C&C) argue that higher-level chunks preserve information from lower-level chunks albeit in a much impoverished form. However, they also suggest that there is no obligatory relationship between low-level chunks and high-level chunks. To support their claim, they cite the case of SF (cf. Ericsson et al. 1980), who could accurately recall as many as 79 digits after grouping them in locally meaningful units (e.g., historical dates and human ages). Moreover, they argue that the Now-or-Never bottleneck forbids broad parallelism in language at the expense of avoiding ambiguities (e.g., “garden path” sentences). In brief, C&C propose that chunks are only locally coherent and that their gist, however contradictory, is being safely kept track of at higher levels. Unfortunately, the authors remain silent about the mechanisms underlying higher-level representation formation.

C&C also declare themselves agnostic about the nature of chunks. Indeed, although there is ample psychological evidence for the existence of chunks in various types of experimental data, from pause durations in reading to naive sentence diagramming, chunks remain notoriously difficult to define. However, we have reasons to reject the possibility, which follows naturally from the Chunk-and-Pass framework, that chunks are arbitrary and may depend exclusively on memory limitations. To wit, chunks correspond most closely to intonational phrases (IPs) (cf. Gee & Grosjean 1983), which, in turn, are hard to capture by grammatical rules. For example, the sentence “This is the cat / that chased the rat / that ate the cheese” contains three IPs (separated by slashes) that fail to correspond to syntactic constituents (noun phrases or verb phrases). Yet IPs are not entirely free of structure, as they must begin at the edge of a syntactic constituent and end before or at the point where a syntactic constituent ends (cf. Jackendoff 2007). Moreover, although a given utterance can be carved up in several ways (hence, contain a variable number of IPs), carvings are not arbitrary and license only certain IP combinations and not others. We may therefore conclude that IPs and corresponding chunks must be globally coherent (i.e., fit well with each other) and depend on the meaning conveyed.

Furthermore, I believe that chunking is driven not by memory limitations nor by language structures, but by an overall need for coherence or meaningfulness (cf. Dumitru 2014). Indeed, evidence from memory enhancement techniques suggests that chunking must rely on global coherence. So, for example, memory contest champions who use the so-called mind palace technique (e.g., Yates 1966) often achieve impressive results. The method requires them to commit to long-term memory a