Individual differences in the real-time neural dynamics of language comprehension

Darren Tanner, Maria Goldshtein, and Benjamin Weissman

Department of Linguistics

University of Illinois at Urbana-Champaign

Contact:

Darren Tanner
Department of Linguistics
University of Illinois
4080 Foreign Languages Building, MC-168
707 S. Mathews Ave.
Urbana, IL 61801 USA
dstanner@gmail.com

Abstract

Recordings of event-related brain potentials (ERPs) provide a rich source of information about the cognitive systems supporting real-time language use. However, the interpretation of ERPs can be complicated by individual differences that aren't reflected in traditional analyses or visualizations. This is problematic, as failure to recognize important and systematic individual differences has in some cases led to inappropriate interpretations of ERP effects, with neurocognitive models of language comprehension sometimes being built on these inappropriate interpretations. In this chapter we review work, largely from our lab, on individual differences in ERP studies of language comprehension and discuss the promise of work on individual differences, as well as the challenges. In some cases individual differences in ERPs manifest themselves quantitatively (i.e., systematic differences in effect amplitudes), but in other more complex cases, qualitatively (i.e., different types of effects in different individuals). We will describe work on individual differences in morphosyntactic and semantic processing in both native and nonnative language processing, as well as multi-modal communication and higher-order pragmatic inferencing. In this last vein we will describe some nascent work done in our lab using unsupervised machine learning algorithms to better understand underlying patterns of qualitative individual differences in the processing of scalar implicatures. We conclude by laying out some challenges and suggestions for future work.

Introduction

Language comprehension is rapid and dynamic, necessitating the incorporation of information from sensory representations (e.g., auditory and/or visual information) with information about grammar, word meaning, and world knowledge stored in long term memory, as well as speaker intent, perspective, and other pragmatic inferences. Event-related brain-potentials (ERPs) provide an excellent measure of these processes as they unfold in real time, as ERPs can provide millisecond-level temporal resolution of cortical electrical activity and can be recorded non-invasively at the scalp. However, traditional ERP visualization in publications uses an average-of-averages method, focusing on effects in grand mean waveforms, and statistical analyses typically use parametric tests for differences in central tendency (e.g., mean amplitude). In very many cases, this averaging process has desirable effects. Indeed, a key facet of ERPs' physiological definition is that they reflect neural electrical signal that is both time- and phase-locked to stimuli of interest (e.g., particular words in sentences, the focus of this chapter), and averaging can extract this signal. Averaging is generally carried out in a two-stage process. First, electrical signals are averaged across trials within experimental conditions within participants to derive subject-level ERPs for each condition. Grand mean ERP waveforms for visualization purposes are derived through the second level of averaging in which the subject-level, condition-specific ERPs are then averaged again across participants. The logic of the averaging process is that it should extract the ERP signal (that is, the electrophysiological activity that is consistently elicited by stimuli in the condition of interest) while the noise (that is, random electrophysiological activity) should approximate zero as the number of stimuli and participants increases. The grand mean waveforms thus depict the central tendency of brain responses across both trials and participants.

In many cases this averaging process has the desired consequences described above of increasing the signal-to-noise ratio and revealing cortical electrical activity tied to processing, and can be very revealing about how language is comprehended in real time. However, there are at least two cases in which the averaging process can miss systematic neural activity observable at the scalp that is tied to dynamic information processing. The first case, which we will not go in depth into here, involves changes in oscillatory neural signals that are time-, but not phase-locked to the eliciting stimulus. In these cases, event-related power changes in various frequency bands in the electroencephalogram (EEG) may be obscured in the averaged ERP because of phase mis-alignment, though they would be detectable in spectrographic analyses of single trial dynamics. Here it is the first, trial-level stage of averaging that would lead to systematic distortion of the signal available in the EEG recording (see, e.g., Bastiaansen, Mazaheri, & Jensen, 2012). The second case involves scenarios where the second stage (between-participant) averaging process leads to inappropriate inferences about processing in the general population. These bogus inferences can arise either when ERP effects are present in grand mean waveforms and reliable in statistical analyses, even when they were not present in most (or even any) of the individuals' ERP waveforms (through topographic or temporal distortion), or when ERP effects are found in some (or many) individuals, but are not reflected in the grand mean or corresponding statistical tests. That is, in some cases, there are meaningful between-participant differences, either quantitative or qualitative, in event-related neural activity that traditional ERP visualization and quantification assume to be noise, but which may in fact be meaningful signal. These individual differences in language processing, with a specific focus on EPRs, are the focus of the present chapter.

Individual differences have not generally been the focus of much the work using ERPs to study language processing. The reasons for this are multi-fold. Although there have always been strong counterpoints, the dominant paradigms in theoretical linguistics in the second half of the twentieth century focused on invariant traits of the human language faculty and focused on commonalities among languages and language users, to the exclusion of variation, variability, and individual differences in grammatical and semantic knowledge. This focus in theoretical linguistics led to a similar narrow focus in more experimental domains of linguistic inquiry. In experimental and cognitive psychology, the dominant experimental paradigms have generally used within-subjects (repeated measures) designs. While repeated measures designs do not necessarily assume total homogeneity of the population being sampled, the focus of this sort of experimental design is not to identify how and/or why individuals differ. Rather, the rationale is to see whether, on average, participant performance (or brain activity) differs between experimenter-manipulated conditions. That is, individual differences have most frequently been relegated to the error term of an ANOVA or another parametric statistical test in most published experimental psycholinguistic investigations, including ERP studies.

Note that this is a common trait that ERP research shares with much work in experimental psychology. Research frequently involves identifying whether an experimental manipulation has an effect on the central tendency of some outcome measure (mean reaction time, mean ERP amplitude, etc.). Inputs to statistical analyses in ERP research, like many other methods, involve either subject-level condition averages or trial-level observations, with parametric tests giving inferences about whether any differences observed in central tendency are statistically reliable (in a traditional null hypothesis testing approach). Grand mean waveforms in ERP research are thus visualizations of the central tendency of amplitude measurements in each condition, as a function of time, much in the same way that tables of descriptive statistics and bar charts provide information about the central tendency of reaction times across conditions in behavioral research. In this way, the outcomes of statistical analyses serve as a confirmation check about whether or not differences seen between grand mean waveforms (that is, central tendency) are reliable. An important way that ERPs differ from measures like reaction time, however, is in their multidimensionality. ERPs are measured in polarity (positive or negative voltage), time, and scalp topography. Distortions due to averaging can impact any combination of these dimensions, making interpretation of effects problematic in some cases. This can be especially tricky in ERP research, as ERP components are generally defined by a combination of these three dimensions. For example, the N400 effect, which is a centro-parietally distributed negativity around 300-500 ms poststimulus, is distinguished from the left anterior negativity (LAN) based solely on scalp topography: the LAN is also a negativity generally occurring between 300-500 ms poststimulus, but primarily over left hemisphere and/or anterior electrodes. As we will describe below, we have shown in several datasets that apparent LAN effects in traditional analyses can be the result superposing a true centro-parietal N400 effect in some participants and a right hemisphere positivity in others (i.e., topographic distortion due to polarity differences across individuals).

Moreover, a common implicit assumption is that the grand mean ERP waveform to a word in a sentence (or more typically a comparison of the ERP elicited by an expected or well-formed word to that elicited by an unexpected or in some way anomalous word) reflects the normative brain response in the population under investigation, be that monolinguals, bilinguals, children, adult second language learners, etc. However, as we will show below, there are many instances where this assumption clearly does not hold. In some cases we find marked variation when studying populations where we would expect marked inter-individual differences (e.g., adult second language learners and bilinguals processing their second language), but we also see variation when studying populations that we previously had assumed to be relatively homogenous in their language processing profiles (i.e., functionally monolingual, literate, right-handed young adults).

As we will explore below, the field has begun to accept that individual differences in the neural dynamics of real-time language comprehension are quite real and are deserving of study. In some cases these individual differences map neatly on to existing hypothesized quantitative individual differences in cognitive or experiential traits relevant to language use, such as verbal working memory capacity (e.g., Kim, Oines, & Miyake, 2018; King & Kutas, 1995; Nakano, Saron, & Swaab, 2010; cf. Just & Carpenter, 1992; King & Just, 1991) or language experience, proficiency, and reading skill (e.g., Mehravari, Emmorey, Prat, Klarman, & Osterhout, 2017; Pakulak & Neville, 2010; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013). Notably, and more interestingly, in many cases in which we have observed clear qualitative individual differences, we have also failed to find clear predictors of ERP response type from among this set of usual candidates, such that there is an immense amount of work left for the future.

The growth of interest in individual differences also comes at a time when the fields of psycholinguistics and language science more broadly are accepting that homogenous monolingualism is not the norm globally, that perfectly balanced biliterate two-natives-in-one bilinguals essentially do not exist, that language use and processing changes with both normal aging and neurocognitive pathology, and that Chomsky’s (1965) idealized speaker-listener is an abstraction that modern psycholinguistics need not (or even cannot) concern itself with. This rising awareness also coincides with an increase in power and availability of computational and statistical tools for understanding variation. Statistical models which previously took hours, days, or even weeks to fit can now be solved within seconds, so an increasing number of models are now within reach of us 'mere mortals.' These include advanced linear models and their variants such as hierarchical/mixed effects linear or generalized linear mixed effects models, and unsupervised learning algorithms that can find hidden structure in datasets. More recently, generalized additive mixed models (GAMMs) have allowed us to relax the assumption of strict linearity in our measures, which is a welcome advance, since ERPs themselves are highly nonlinear (they show positive and negative voltage changes over time), and ERPs (or other measures of processing) may have non-linear associations with individual difference measures of interest across either time or space. Some nascent work has begun to apply GAMMs to ERP individual difference problems (e.g., Meulman, Wieling, Sprenger, Stowe, & Schmid, 2015; Tremblay & Newman, 2015), though the computational challenges involved in applying GAMMs to ERPs' multiple dimensions and participant-level individual differences simultaneously remain considerable. Note also that the fact that individual differences, and in particular qualitative individual differences, in ERPs have only recently begun to be systematically explored means that we have only just begun to understand the dimensions in which variation in ERPs manifests itself, and how to quantify variation across these dimensions. For these reasons the scope of this chapter will necessarily be limited, and it is intended as a starting point for future work on this issue.

In the next section we will briefly review some of the most commonly studied ERP components related to language processing and highlight a few of the foundational studies of individual differences that have inspired much of our own work. In the subsequent sections we will review some work on individual differences using ERPs, some from other groups and some from our own, some older, and some new and emerging, which looks at individual differences in ERPs in multiple domains: non-native processing and bilingualism, morphosyntactic and semantic processing in monolingualism, multi-modal communication and pragmatic inferencing, and some nascent approaches to unsupervised machine learning to classify sub-groups of participants based on brain responses in an unbiased way. We will finish with some concluding remarks about challenges and suggestions for the future.

Language-related ERP Components

By far, the two most commonly studied ERP components in studies of sentence and discourse comprehension are the N400 and the P600, with the N400 being the better-understood component of the two. Since first being reported in 1980 (Kutas & Hillyard, 1980; see Kutas & Federmeier, 2011, for an extensive overview) the N400 has been understood to broadly reflect semantic processing, with many interpretations holding that the amplitude of the N400 indexes degree of difficulty accessing meaning from long term semantic memory (Kutas & Federmeier, 2000, 2011; Lau, Phillips, & Poeppel, 2008; though see Hagoort, Baggio, & Willems, 2009): larger amplitude N400s reflect more difficulty accessing a word's semantics due to the word being unexpected or anomalous, and smaller amplitude N400s reflect facilitated semantic access due to strong contextual support, priming, and/or (partial) preactivation of a word or aspects of a word's meaning.

The P600, on the other hand, was first discovered in the context of syntactic anomalies: dis-preferred garden path continuations, subcategorization violations, and morphosyntactic (agreement) anomalies (Hagoort, Brown, & Groothusen, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992, 1993). Later work found that the P600 could be elicited in other syntactically well-formed, but difficult to process, constructions, such as WH-dependencies, where small P600 effects were elicited at the gap position (Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005), making the interpretation of the P600 as an index of syntactic anomaly difficult. Later in the 2000s P600s were reported in sentences that were grammatically simple and well-formed, but which violated some semantic constraint (e.g., verb-argument animacy constraints), or contained thematic role reversals, or simply contained very strong implausibilities (as opposed to mild implausibilities where only N400 effects were elicited; Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010; van Herten, Chwilla, & Kolk, 2006). Additional studies have reported that the P600 could be reduced or eliminated when the sentence contained misleading cues about morphosyntactic relations that made unambiguous syntactic violations less salient and/or when words were presented at rates faster than those typically used in ERP studies (Tanner, Grey, & Van Hell, 2017; Tanner, Nicol, & Brehm, 2014). Still others have found that anomalies involving structured non-linguistic representations (e.g., music and visual events) can also elicit P600 effects (Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008). This disparate set of eliciting conditions has challenged the notion that the P600 reflects primarily syntactic processing or syntactic reanalysis, and a number of non-mutually-exclusive views of the P600's functional significance have been put forth over the last decade (see e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Brouwer, Fitz, & Hoeks, 2012; Kolk & Chwilla, 2007; Kuperberg, 2007; Paczynski & Kuperberg, 2012). Nonetheless, the N400/P600 dichotomy has played a pivotal role in framing much psycholinguistic ERP work in both monolingual and bilingual contexts over the last three decades, and as such, it will play a key role in the present review as well.

Although the N400 and P600 have occupied focal roles in ERP work on sentence processing, one additional component deserves mention here. In some reports, morphosyntactic violations elicit a negative-going wave that sometimes has a left and/or anterior scalp topography preceding the P600: the left anterior negativity, or LAN effect. The LAN surfaces in the same time window as the N400 effect (around 300-500 ms poststimulus), but is distinguished based on its scalp topography: the N400 typically has a broad central or centro-parietal topography, whereas the LAN has the aforementioned left and/or anterior distribution. The LAN features prominently in some neurocognitive theories of morphosyntactic processing as an index of initial detection of anomalies (which in some theories is additionally referred to as an automatic, implicit process), where later controlled reanalysis is indexed by the P600 effect (e.g., Batterink & Neville, 2013; Molinaro, Barber, Caffarra, & Carreiras, 2015; Molinaro, Barber, & Carreiras, 2011; Ullman, 2004), though there are numerous conceptual and empirical issues with these claims, some of which we will discuss below (see also Osterhout, McLaughlin, Kim, Greewald, & Inoue, 2004; Tanner & Van Hell, 2014, for further discussion).

It is important to note here that the N400, P600, and LAN are not language-specific, nor are they the only ERP components studied in language-related electrophysiological research. However, they are large, and, at least in the case of the N400 and P600, easily elicited by sentence-embedded linguistic manipulations. They have been the focus of the majority of research on native and non-native sentence comprehension. See work by Swaab and colleagues (Swaab, Ledoux, Camblin, & Boudewyn, 2012) and Payne and colleagues (Payne, Ng, Shantz, & Federmeier, in press) for recent reviews of ERP research on language.

Some Quantitative Individual Differences

The most straightforward way in which individual differences in ERPs may manifest themselves is quantitatively. When we say "quantitatively," we mean that the individual difference 'signal' (that is, the inter-individual variability that is systematic and quantifiable) is found in some component's effect amplitude, e.g., the N400 or P600, and we can find some interesting regressor that predicts the magnitude of the effect across individuals. Here effect amplitude will typically refer to the difference in amplitude between some ill-formed and well-formed condition (e.g., P600 effect amplitude will be the mean amplitude difference in the P600 time window over posterior electrodes from the ungrammatical minus grammatical difference wave in a morphosyntactic violation paradigm).

One notable example that challenged the notion that monolinguals are homogenous is reported by Pakulak and Neville (2010). They investigated syntactic processing in monolingual English speakers, but used a relatively large sample size compared to most ERP studies (N = 72) and sampled not just university-enrolled subject pool participants that are typically tested in ERP studies, but also tested participants with a range of literacy and socioeconomic status levels. They recorded ERPs to auditorily presented sentences that were either well-formed or included phrase structure violations (*Timmy can ride the horse at* <u>*his*</u> *farm.*/**Timmy can ride the horse at my* <u>*his*</u> *farm.*), and also administered a very difficult English proficiency test to their participants (the Test of Adolescent and Adult Language, 3rd Edition, TOAL-3). ERPs elicited by the well-formed and anomalous determiners (underlined, above) showed an early LAN and P600 in the grand mean analysis; however, they also found individual differences that varied with participants' proficiency. In particular they found that the LAN was more bilateral and the P600 of smaller amplitude in lower proficiency participants, after controlling for working memory capacity. This finding is notable for two reasons. First, as mentioned above, it provided a clear counterpoint to the notion that all native speakers are equally proficient in their native language and that processing resources are homogenous across monolingual speaker populations. Second, it showed quite clearly that between-subjects variation in ERP measures can be systematic, such that – especially in the case of the LAN topography – traditional omnibus statistical analyses of mean amplitudes (as would be depicted in grand mean waveforms) may not provide an accurate depiction of ERP effects in many of the individuals in the sample under investigation. More generally, this study set the stage for the many psycholinguistic ERP studies on individual differences in the subsequent years.

With Pakulak and Neville having shown that language proficiency, over and above working memory, can account for quantitative differences in ERP metrics of syntactic processing, we asked whether a similar observation might hold for semantic processing and the N400. Some have argued that working memory skill is particularly important in semantic processing, as managing competing demands of accessing words and building message level meaning can be cognitively taxing (e.g., Boudewyn, Long, & Swaab, 2013). Others have argued specifically that working memory is less important in sentence comprehension, and that semantic activation and retrieval is mediated by "lexical quality," which is a function of language experience and proficiency (e.g., Van Dyke, Johns, & Kukona, 2014; cf. Perfetti, 2007). In a preliminary report, we evaluated these claims in a large cohort of monolingual English speakers (N = 64; Tanner, Bulkes, Shantz, Armstrong, & Reyes, 2016) by recording ERPs while participants read sentences that were well-formed or contained a word that was semantically unexpected and anomalous (e.g., *The doctor diagnosed the <u>tumor/*package</u> while it was still* *treatable; The delivery man left the package/*tumor on the front porch*; the underlined word marks the time-locking point for ERP averaging and analysis). We also administered a battery of behavioral tasks tapping language experience and lexical knowledge, as well as working memory. The semantically anomalous words elicited much larger amplitude N400 effects than their corresponding well-formed counterparts (the N400 effect), as predicted. However, our question was whether the individual differences related to working memory and/or language experience would predict the N400 effect amplitudes. Because we administered a battery of tasks, we used principal components analysis (PCA) to extract component scores for two latent components; the component loadings showed that principal component (PC) 1 correlated strongly with the language experience tasks and PC 2 correlated strongly with the working memory tasks. We then regressed these latent component scores onto N400 effect amplitudes simultaneously. Scatterplots of the zero-order correlations are depicted in Figure 1. There was no correlation between the latent language experience component and N400 amplitude (r = -.006), but there was a reliable correlation between participants' latent working memory component scores and N400 effect amplitudes (r = .325, p = .009).

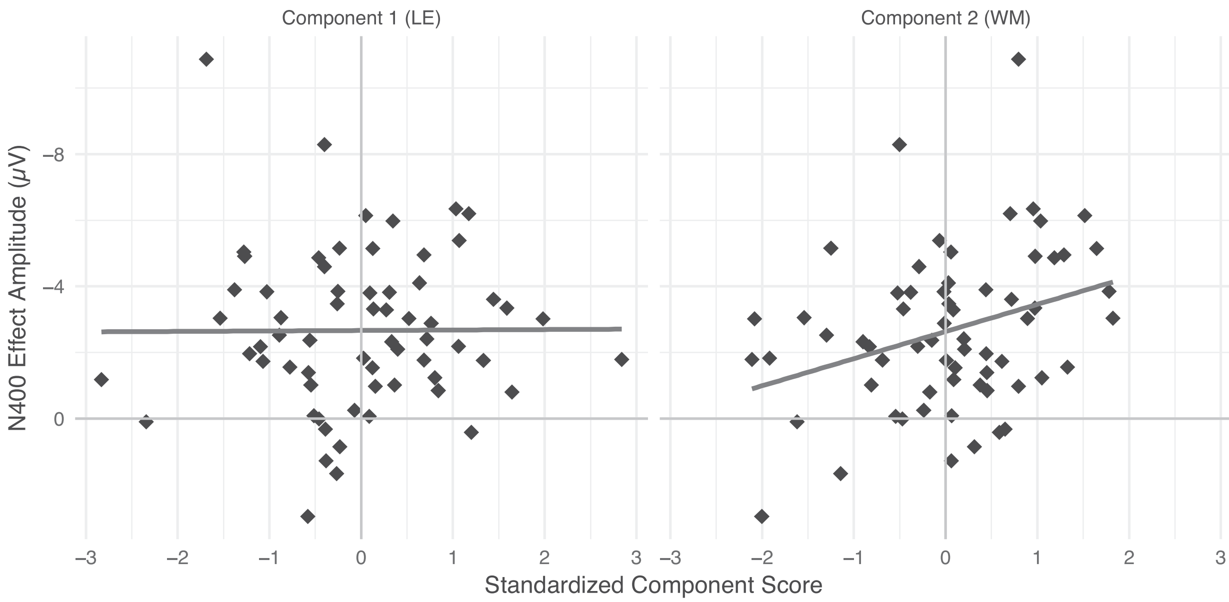

Figure 1. Scatterplot depicting relationship between latent component scores for language experience (left) and working memory (right) and N400 effect amplitude (y-axis) for 64 monolingual English speakers, measured across eight centro-parietal electrodes.

These findings show that, among cognitive skills commonly implicated in language processing, there may be some kind of dissociation between individual differences supporting semantic and syntactic processing. Pakulak and Neville (2010) reported that a speaker's proficiency in their native language, over and above working memory, was predictive of P600 amplitudes in syntactic violation contexts; we found that speakers' working memory capacities, over and above native language proficiency, were more predictive of N400 effects in semantic violation contexts. Note that firm conclusions cannot be drawn from this comparison, since the modalities of the two studies were different (auditory versus written presentation), the types of participants sampled were different (a broad spectrum of literacy levels versus literate young adults, who nonetheless varied in their native language proficiency), and the metrics used to assess working memory and language knowledge differed across the two studies. A clear test of this hypothesis should use a within-subjects design, and this remains an active area of work within our lab. Nonetheless, these two studies demonstrate that grand mean analyses can obscure some systematic variation between individuals in ERP responses. In the present scenarios the mappings between ERP responses and behavioral predictors was relatively straight-forward: ERP effect amplitudes in an expected elicited condition (N400 for semantic violations and P600 for syntactic violations) showed linear associations with some well-established behavioral measure. However, as we will see in the coming sections, such straight-forward mappings are not ubiquitous, and the relationship between expected and observed ERPs can be quite complex.

Qualitative Individual Differences in Non-native and Bilingual Morphosyntactic Processing

Our initial work on qualitative individual differences ERP responses was inspired by two lines of research laid out by Lee Osterhout and colleagues. First, in a series of review papers, Osterhout, Judith McLaughlin, and colleagues reported findings from a longitudinal study of second language (L2) learning, which tracked native (L1) English speakers as they progressed through their first year of classroom L2 French instruction (McLaughlin et al., 2010; Osterhout et al., 2008, 2004; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006). Most relevant for us, they reported qualitative changes in participants' ERP responses to morphosyntactic subject-verb agreement violations over the course of the year: violations, which elicited P600 effects in native French speakers, initially elicited N400 effects in the L2 group, but the L2 participants' brain responses changed over the course of instruction to approximate natives' P600 by the end of the year. That is, the L2 learners initially processed the agreement violations as though they were lexical or semantic anomalies, and their brain responses changed qualitatively. Second, a separate line of work by Osterhout (1997) in monolingual English speakers demonstrated that a biphasic LAN-like negativity followed by a P600 in grand mean ERPs elicited by garden path disambiguations was not representative of most participants' brain responses. Rather, most participants showed either an N400-like negativity or a P600, but not both, and that the scalp topography of the negativity in the grand mean showed an anterior focus only because the posterior-dominant P600 effect cancelled out the posterior portion of the broad N400 component due to component overlap.

These two threads helped us better understand a perplexing pattern of data found in native English speakers with low proficiency in L2 German, processing German morphosyntactic agreement violations (Tanner et al., 2013). Unlike native and intermediate proficiency L2 German speakers who showed P600 effects, grand mean ERP waveforms in the group of 20 first year German students showed a biphasic N400-P600 response (Figure 2). However, inspection of individuals' ERPs showed that this pattern was not representative of most participants' brain responses. The distribution of individuals' brain responses is shown in Figure 2B. Individuals' N400 effect amplitudes are the y-axis (larger N400 effects are up on the graph) and P600 effect amplitudes are on the x-axis (larger P600 effects are to the right). Across participants, the two ERP responses showed a negative association: individuals who showed large N400 effects tended to show small P600 effects, and vice versa. The dashed line indicates the axis of equal effect sizes. Individuals above/to the left of the line show N400-dominant ERP responses and individuals below/to the right of the line show P600-dominant ERP responses. Averaged waveforms within the N400- and P600-dominant groups (Figure 2C and D) showed clear monophasic responses of the respective type. Subsequent correlation analyses showed that participants' behavioral discrimination of the well-formed and ill-formed sentences correlated negatively with N400 effect amplitudes and positively with P600 effect amplitudes, suggesting that individual differences in proficiency were partly responsible for the pattern of qualitative individual differences in ERP responses. In many ways this finding is a cross-sectional replication of Osterhout and McLaughlin's original longitudinal study, but with the qualitative change in ERP responses associated with proficiency being captured in a single "snapshot." Importantly, this finding shows that the grand mean waveforms and corresponding omnibus statistical analysis misrepresented the true nature of many participants' brain responses. Even though the first-year learners had relatively similar and homogenous classroom exposure to L2 German, their processing differed from one another in marked, yet systematic ways.

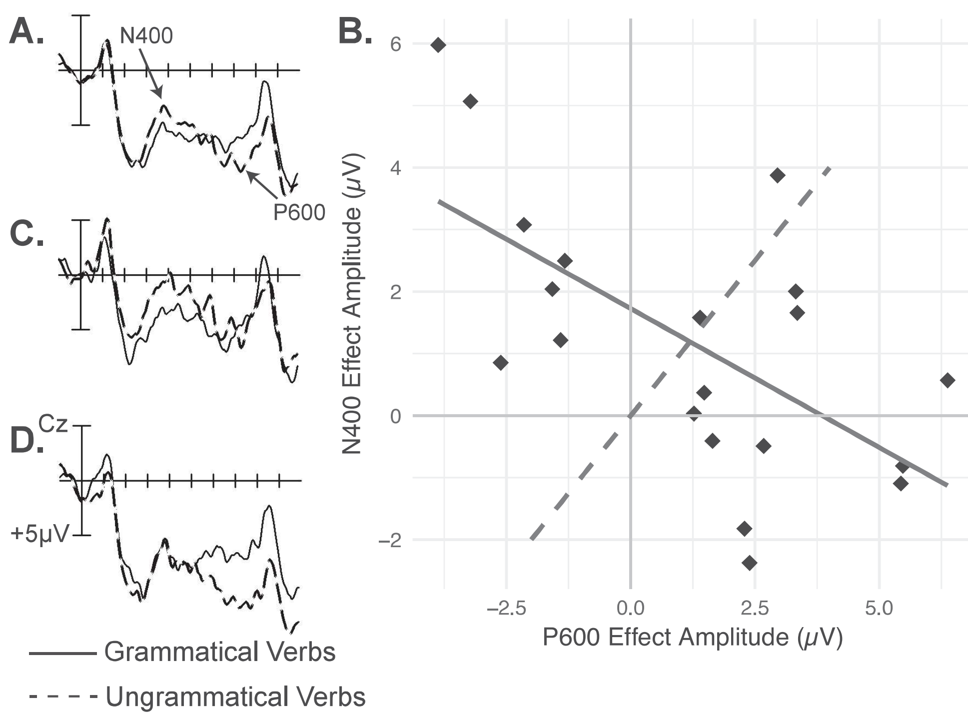

Figure 2. Partial results from Tanner et al. (2013). (A) Grand mean ERPs from 20 L1 English speakers enrolled in first year L2 German courses to well-formed control (solid) and subject-verb agreement violating verbs (dashed). Negative voltage is plotted up in these and all subsequent waveforms. (B) Distribution of the first year participants' N400 and P600 ERP effect amplitudes. (C) ERP responses in the negativity-dominant group (n = 9). (D) ERP responses in the positivity-dominant group (n = 11).

In a subsequent study (Tanner, Inoue, & Osterhout, 2014) we investigated morphosyntactic processing in very high proficiency late L1 Spanish-L2 English bilinguals and found a similar distribution of brain responses between N400- and P600-dominant in this group in response to subject-verb agreement violations in English. This was actually quite surprising to us, since previous work had found that ERP responses in non-native language processing could approximate natives' quickly for shared morphosyntactic features, and we specifically targeted high proficiency bilinguals for this study. Nonetheless, English subject-verb agreement violations elicited N400 effects in some, biphasic responses in some, and P600 effects in others. To better understand what factors were related to these qualitative individual differences, we created two new measures for the types of variation we found. The Response Dominance Index (RDI) is a metric of how N400- or P600-dominant a participant's brain response is. Negative values reflect N400-dominance, positive values reflect P600-dominance, and values near zero reflect dominance of neither one response nor the other. However, the RDI does not directly index the *size* of the response, only the relative size of the polarity dominance. In principle, participants with an RDI of zero could show a large, but equisized biphasic response. To complement the RDI, we therefore also proposed the Response Magnitude Index (RMI), which provides an index of the total "amount" of brain response, measured from the anomalous minus control difference wave, regardless of brain response type. RMI values near zero reflect little to no difference in brain responses between the anomaly and control conditions, and large values reflect greater neural sensitivity to the anomalies, but without indicating the identity of the ERP component elicited.

To identify the antecedents of the individual variation in ERP responses among the bilinguals we tested, we computed two multiple regression models, with the RDI and RMI as the respective dependent measures. Predictors included age of arrival in the United States, length of residence in the US, frequency of use of English in daily life, English proficiency (as measured via an abbreviated version of the Michigan Examination for the Certificate of Proficiency in

English), and self-rated motivation to speak English like a native speaker. Results showed that RDI values were predicted by age of arrival and motivation, with more P600 bias found among those who arrived earlier and who reported higher motivation to speak like a native speaker. RMI values were predicted by proficiency: higher proficiency correlated uniquely with greater neural sensitivity to subject-verb agreement violations (but not a specific ERP component). For the age of arrival effect, note that all of the participants in this group were late, post-puberty arrivals, putting their arrival in the US after any putative critical period cut-off. The correlation between age and ERP response quality therefore does not lend itself to a straight-forward Critical Period Hypothesis explanation, but is consistent with age effects on L2 learning and processing, broadly defined (cf. Birdsong, 2006; Birdsong & Molis, 2001).

The fact that substantial individual variation existed among the highly proficient bilinguals tested in this second study on the one hand may not be surprising, since the circumstances of bilingualism and L2 learning are highly heterogeneous. However, there are several proposals about L2 processing holding that neural markers of non-native processing should converge on "native-like" ERP responses at high L2 proficiency (i.e., a P600 effect to subject-verb agreement violations; see Morgan-Short, 2014; Steinhauer, 2014, for reviews). One might infer, then, that the participants who showed N400 effects in response to English subject-verb agreement violations were poorer learners or processors, since they showed non-native-like ERPs, which were previously shown to be associated with low proficiency L2 learners. However, this inference rests upon the assumption that monolinguals show homogeneous brain responses in their L1 and that no proficient monolinguals show N400 effects. A cursory reading of the literature would lead one to believe that this is the case; however, as we discuss below, careful study shows that even proficient, literate monolinguals can differ qualitatively from one another in their ERP responses, and a substantial minority indeed show N400 effects to morphosyntactic violations.

Monolingual Morphosyntactic, Semantic, and Thematic Processing and Some Psycholinguistic Conundrums

In Tanner and Van Hell (2014a) we sought to identify whether similar qualitative individual variation exists among proficient, literate monolingual English speakers processing morphosyntactic relations, and in particular, whether individual variability could explain at least some of the unreliability of the LAN component across studies. Although the LAN features prominently in some ERP-based neurocognitive theories of language processing, including some arguments that it reflects automatic syntactic processing or anomaly detection, a substantial proportion of ERP studies fail to find a LAN in response to morphosyntactic violations (meaning, by definition, that the effect can therefore not be strictly automatic). Based on findings from Osterhout (1997) and our work on individual differences in L2 processing, we hypothesized that at least some instances of biphasic LAN-P600 effects in the reported literature were artifacts of qualitative individual differences, wherein some participants showed N400 effects and some showed P600 effects. Indeed, this is what we found. In response to morphosyntactic violations, grand mean waveforms and traditional omnibus analyses showed a statistically robust left hemisphere (LAN-like) negativity followed by a P600 effect. In this case the LAN was found to be statistically reliable via a grammaticality by hemisphere interaction over lateral electrode sites from both hemispheres in the 300-500 ms time window, i.e., without isolating left hemisphere electrodes alone as a region of interest. However, further investigation revealed a continuous distribution of brain responses between N400- and P600-dominant, very similar to that seen in our previous work on non-native processing. Analyses of the semi-homogeneous sub-groups (negativity- and positivity-dominant groups, respectively) showed that the negativity was not left lateralized, but instead showed a broad, N400-like centro-parietal distribution. The P600 in the other group, on the other hand, showed a right-posterior dominance, with an onset during the 300-500 ms time window over right hemisphere electrodes. That is, the LAN that was detected in the grand mean was a consequence of component overlap and individual differences: the left hemisphere topography was a result of the intersection of a broadly distributed N400 in some individuals and a temporally co-occurring right hemisphere positivity in others, with the residual N400 being visible only over left hemisphere electrodes in the grand mean. This gave the illusion of a LAN effect. This negativity was still statistically significant in omnibus tests of central tendency, despite being present only in a minority of participants. This suggests that at least some of the variation across studies in the LAN might be explained by individual differences and sampling variability; i.e., the presence or absence of a LAN seen in grand mean analyses would be at least partly a function of the number of N400-dominant individuals randomly sampled for a study, as well as the degree of spatiotemporal overlap between the P600 seen in most individuals and the N400 seen in others.

On the one hand, this result suggests that the N400 effects seen in some non-native speakers cannot be clearly interpreted as evidence of non-native-like processing or poor L2 learning, since a substantial proportion of native speakers also showed N400 effects in similar morphosyntactic situations when the data were more closely inspected. However, given that L2-related regressors predicted some proportion of variance in brain response dominance in the previous work (acceptability judgment accuracy as a proxy for L2 proficiency, age of arrival, and motivation to speak in a native-like way), the question remains what individual-level metrics might predict brain response dominance in native speakers. In the Tanner and Van Hell (2014a) study, familial sinistrality accounted for a small proportion of the variance in a way predicted by Bever and colleagues: right-handed participants with left-handed family members showed, on average, more negative-going ERP responses than right-handers without left-handed family members (Hancock & Bever, 2013; Townsend, Carrithers, & Bever, 2001; see also Lee & Federmeier, 2015). A follow-up on this study with left-handed individuals corroborated this finding by showing that left-handers and right-handers with left-handed family members patterned similarly, and both groups had more variation in the response dominance continuum than the right-hander group without familial sinistrality (Grey, Tanner, & Van Hell, 2017). Our lab has also recently completed a large-scale (N = 114) study of individual differences in morphosyntactic processing in English monolinguals where we administered a battery of language experience and working memory measures, in combination with latent variable analyses of language experience and working memory, in an effort to identify cognitive antecedents of variation in the N400-P600 response dominance continuum (Tanner, 2018). This study also addressed some criticisms leveled by Molinaro and colleagues (Molinaro et al., 2015) about the syntactic complexity of the materials used in the Tanner and Van Hell study. Using structurally simple sentences, traditional analyses showed a similar LAN-P600 pattern to subject-verb agreement violations, but again, there were marked individual differences in the N400-P600 continuum, both for lexical verbs where agreement is marked with an overt morpheme (*grow/grows*) and for auxiliary verbs where agreement is marked by lexical alternations (*is/are*). Again, the apparent LAN in the grand mean was a consequence of component overlap between an N400 effect in some individuals and right hemisphere P600 effects in others. However, latent variables related to neither working memory nor language experience were correlated with individuals' ERP effect amplitudes. This is despite findings of robust effects of language experience/proficiency on the P600 effect in our L2 work and in monolingual work by Pakulak and Neville (2010), and despite findings of working memory effects in predicting ERP response dominance in contexts of animacy violations (Kim et al., 2018; Nakano et al., 2010). Thus, although the finding of N400/P600 individual differences in monolinguals is robust and replicable, we do not yet have a clear picture regarding what individual-level traits might predict these differences.

As individual differences helped explain a conundrum regarding variability in the LAN, we asked whether they could help explain conflicting findings in the literature regarding the role of semantic associations between words and their interaction with animacy constraints in thematic role processing. Some have reported that verb-argument animacy violations elicit P600 effects regardless of the verb's semantic association with its subject argument (e.g., Kuperberg et al., 2006, 2007). Others have reported that semantic associations are evaluated independently of combinatorial constraints like animacy and morphosyntax, such that animacy violations involving semantically associated verbs and arguments elicit P600 effects (so-called "semantic attraction"), while animacy violations with unrelated verbs and arguments elicit N400 effects (e.g., Kim et al., 2018; Kim & Osterhout, 2005). We assessed the impacts of morphosyntactic well-formedness and semantic association during thematic role processing in a cohort of 58 monolingual English speakers through an individual differences lens (Tanner & Van Hell, in prep; Tanner & Van Hell, 2014b). The control condition included passive sentences with inanimate subject noun phrases, which were highly semantically related to the verb (*The broken television was <u>repaired</u>...*). Experimental sentences included a syntactic violation condition (*The broken television was <u>repairs</u>...*), a condition with animacy violations indicated by progressive verb morphology and where there was no semantic association between the subject and verb (semantic violation condition: *The hearty meal was repairing*...), and an animacy violation with strong semantic attraction between the subject and verb (the attraction violation condition: *The broken television was repairing*...).

Results are depicted in Figure 3. Grand mean results in the syntactic violation condition showed a large, biphasic N400-P600 effect, a clear N400 effect in the semantic violation condition, and a small, marginal biphasic N400-P600 in the attraction violation condition. However, as with the above studies, interpreting the ERP effects seen in the grand mean waveforms – and confirmed via parametric statistical tests – as representing the normative brain responses in the general population would have led to the wrong conclusions. Instead, the biphasic responses were the result of individual differences in N400/P600 response dominance (Figure 3B). Additionally, the monophasic N400 effect in the semantic (no-attraction animacy) violation condition misrepresented the fact that some individuals showed monophasic positive responses in this condition (bottom right quadrants of scatterplots in Figure 3B). Although there were individual differences in all three conditions, the central tendency of the distribution of differences for the syntactic and semantic violation conditions was moved in the direction predicted by the "classic" take on ERP responses to syntactic and semantic violations: the RDI showed a positive mean in the syntactic condition, indicating a preponderance of P600 effects, and a negative mean in the semantic condition, indicating a preponderance of N400 effects. In the attraction violation condition, there was no clear bias toward positive or negative effects, with the distribution approximately equally distributed between the two effect dominances (Figure 3C). Participants' RDI values across the three experimental conditions were positively correlated (rs > .5), indicating some consistency in response type (N400 versus P600 dominant) within individuals across the various anomaly types. Also interesting to note is that, even though the ERP response visible in the grand mean to attraction violations was small and only marginally significant, the mean of individuals' overall response magnitudes was similar to that seen in the semantic violation condition (Figure 3C). The effect amplitudes in measures of central tendency were attenuated due to cross-participant cancellation between co-temporal negativities and positivities: the P600 showed an early onset in some individuals and cancelled out the N400 seen in others, and the N400 showed a long time course in some individuals, effectively cancelling out the P600 seen in others.

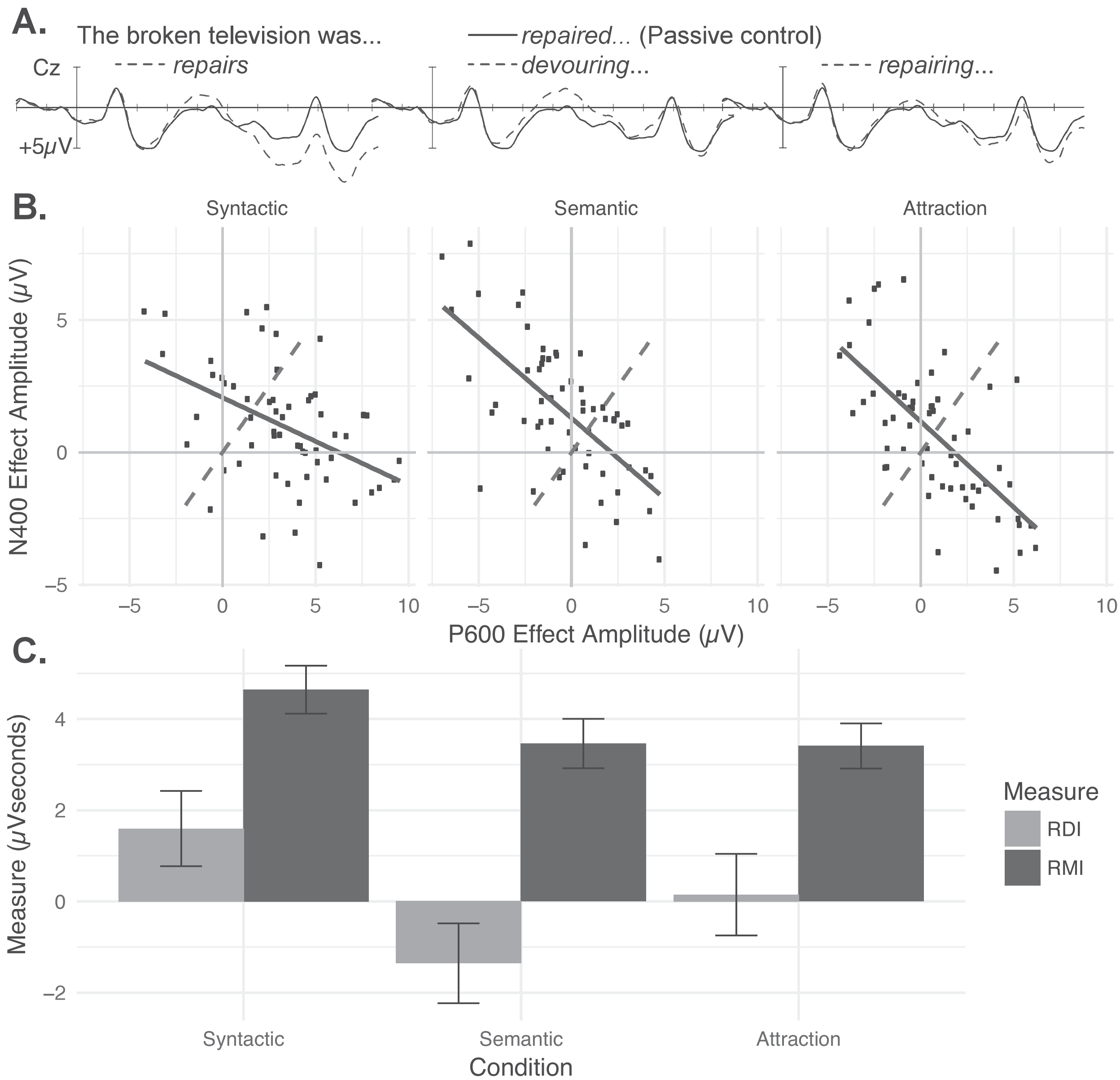

Figure 3. Results from Tanner and Van Hell (2014b). (A) Grand mean ERP results (N = 58) in the three experimental conditions, relative to passive control sentences. (B) Individual differences in N400/P600 effect amplitudes. (C) Measures of ERP response dominance (RDI) and response magnitude (RMI) in the three experimental conditions. RDI measured as positive minus negative integration areas under individuals' difference waves between 200 and 1000 ms poststimulus; RMI measured as rectified area under difference wave. Error bars show 95% CI of the mean.

This last study shows that, although ERP responses to a range of violations (morphosyntactic, semantic, thematic) show individual differences in the N400/P600 continuum, the distribution of these differences in sensitive to the amount and type of information signaling the anomaly in a way that reflects traditional interpretations of the N400 and P600 effects. ERPs were most positive-going when morphosyntactic information signaled an anomaly, most negative-going when violations involved incoherent semantic relations, and there was no clear dominance (equal distribution between N400 and P600 effects) when both semantic and morphosyntactic information were available to signal violations (i.e., world knowledge about what broken televisions can do, and morphological marking on the verb signaling thematic role assignment).

Higher-order Inferencing and Multi-modal Communication

Beyond communicating simple morphosyntactic, semantic or thematic relations, an interlocuter must also consider the speaker's intent, and must sometimes compute semantic interpretations that go beyond the literal meaning of the words used in the utterance. These phenomena, of which there are many, can be broadly classified as "pragmatic inferences." Pragmatics, with its rich theoretical foundation, has turned increasingly often to experimental methods within the past two decades (see e.g., Noveck & Sperber, 2004) in order to better understand the time course and processing of these inferences. In this section we discuss two ERP studies that explore individual differences regarding two types of pragmatic phenomena: scalar implicature and multi-modal irony.

Scalar implicatures are generated by interlocutors when they hear a scalar term like *some* and they have reasons (e.g. context, previous experience with this term) to assume the speaker chose *some* over a stronger term (like *all*) due to the stronger term being false or not applicable. When a speaker says *I ate some cookies*, an interlocutor would often interpret that to mean that the speaker did not eat *all* cookies (in contexts where *all* is a relevant alternative). Underinformative *some* statements like *Some people have lungs* have often been used to investigate whether the literal meaning of *some* ("and possibly all") is generated prior to the pragmatically enriched meaning ("but not all"). A common finding is that participants divide into two camps (of varying sizes, depending on the experimental design, the alternative quantifiers used in the experiment, and more): those who judge underinformative *some* items as true, and those who judge them as false (Bott & Noveck, 2004; Degen & Tanenhaus, 2016; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2009; Papafragou & Musolino, 2003). Several studies have attempted to characterize the two types of responders, the more 'logical' ones, described as relying on the literal meaning, and the more 'pragmatic' ones who rely on the pragmatically enriched meaning. Noveck & Posada (2003) measured truth-value judgments and ERP responses to items like *Some sentences have words* and found no difference in ERP response for the two types of responders, though there were reaction time differences between the two groups.

More recently, others have focused on whether autistic-like traits might underlie differences in processing. Some previous studies have shown participants with autism spectrum disorder (ASD) to be worse than neurotypical populations at understanding metaphor, irony, sarcasm and jokes (Happé, 1993, 1994; Martin & Mcdonald, 2004). With respect to scalar implicatures, ASD participants have been shown to have difficulties in attributing mental states to others (Baron-Cohen, 2000), which directly relates to interpreting communicative intentions (Sperber & Wilson, 1986) and with making inferences related to world-knowledge (Jolliffe & Baron-Cohen, 1999, 2000). For example, Nieuwland, Ditman, and Kuperberg (2010) correlated individuals' Autism Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley,

2001) scores with their ERP responses elicited by temporarily underinformative scalar implicature sentences, compared to sentence with informative scalar implicatures (e.g., *Some people have pets/*lungs, which require good care.*). They found that individuals with more autistic-like traits showed reduced N400 effects to underinformative compared to informative *some* items, whereas participants with less autistic-like traits showed larger N400 effects to the same items. However, their design did not incorporate truth value judgments, so it is not clear how these ERP differences might map onto differences in interpretation. Some more recent behavioral work has found that judgments of underinformative *some* items did not correlate with participants' AQ scores (Antoniou, Cummins, & Katsos, 2016). In another ERP study with a visual verification task, Spychalska, Kontinen, and Werning (2016) found that participants who judged underinformative items as false (pragmatic responders) showed a biphasic N400-P600 effect to underinformative compared to true items, whereas those who judged underinformative items as true (logical responders) showed no ERP differences in the underinformative versus true conditions.

The difference in the methodologies and stimuli used in these studies makes it difficult to determine whether people differ in how they process underinformative *some* items, whether these differences map onto different judgments, and what traits (e.g., ASD-like traits) modulate the differences in judgments and ERP responses. In an attempt to triangulate the different hypotheses about individual differences in the processing and judgment of underinformative *some* items within a single design, we have consolidated some of the previous studies' methodologies and stimuli into one paradigm. Sixty native English speakers read sentences of the form *Some x are y* that were either true (*Some animals are mammals*), false (*Some reptiles are mammals*), or underinformative (*Some cats are mammals*), followed by a true/false judgment. After the ERP experiment, participants completed the AQ-39 questionnaire (Lau et al., 2013) to measure their degree of autistic-like traits (higher score = more autistic-like traits). Results (Figure 4A) showed that, relative to final words in true items, final words in false items elicited larger N400 amplitudes, and final words in underinformative items elicited a reduced N400 followed by a later posterior P600-like positivity. However, these responses were not reliably mediated by participants' AQ scores, as had been reported by Nieuwland et al. (2010; see Figure 4B and C). Correlations between participants' AQ scores and N400 effect amplitudes (following Nieuwland et al.) were in the same direction as in Nieuwland et al.'s study, but did not reach significance, despite a larger sample size (r = -.16, p = .22). Although our items and task differed from Nieuwland et al.'s, our results suggest that the relationship between autistic traits and ERP effects elicited by underinformative items may not be as strong or as systematic as some studies may imply. Note also that we found very little variation in participants' truth value judgments, with nearly all participants providing "logical"-type responses. We therefore could not directly investigate the relationship between truth value judgments and ERP responses or autistic traits. However, we nonetheless found a difference in ERPs to underinformative versus true items among our predominantly logical-responding participants, in contrast to Spychalska and colleagues.

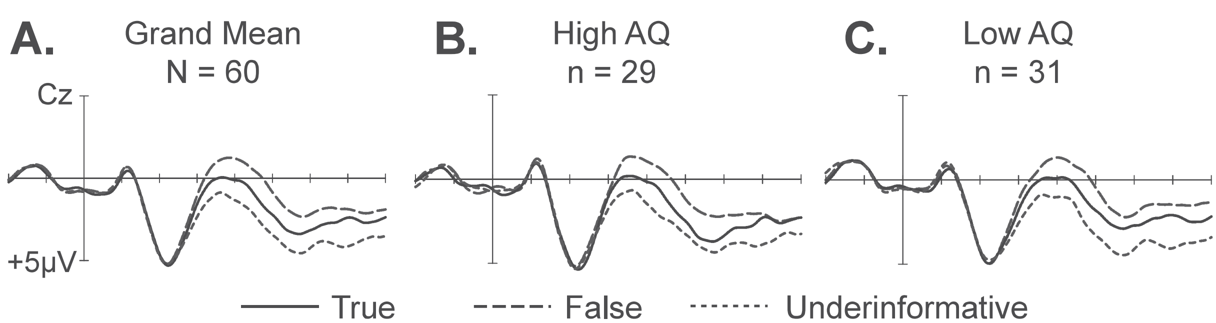

Figure 4. (A) Grand mean ERPs. (B) and (C) Sub-group averages based on median split for AQ: high AQ (B) and low AQ (C). Higher AQ scores reflect more autistic-like traits.

Although we could not establish a clear connection between participants' ERPs and degree of autistic-like traits, we nonetheless observed that individuals' responses to the target conditions did differ qualitatively. It seemed possible, then, that the grand mean waveforms and omnibus analysis obscured variation that was not accounted for by the AQ variable. In order to better understand this variation in a principled, bottom-up way, we have conducted a series of exploratory analyses using unsupervised clustering algorithms to identify whether there are semi-homogeneous sub-groups. One of these analyses used the k-means algorithm to identify clusters for both the false and underinformative conditions, relative to the true (control) condition. K-means clustering is an algorithm that partitions observations into an assigned number of clusters (k). In each run of the algorithm, observations are assigned to a cluster with the nearest centroid mean. This is done iteratively until cluster assignments are stable. We tested models for both the false and underinformative conditions with between two and ten clusters. Model inspection showed that either two or three clusters provided optimally parsimonious solutions for both conditions; however, for brevity we report only the two-cluster solutions here.

Clustering results for the false and underinformative conditions (relative to true) are shown in Figure 5 and Figure 6, respectively. For the false/true contrast, one cluster showed the clear enhanced N400 effect for false items relative to true seen in the grand mean. The negativity showed a long time course (beyond the typical 300-500 ms time window), and became more frontally distributed in the later 500-800 ms time window. The second cluster showed no clear differences between conditions in the N400 time window, but instead a small positivity in the later time window with a right hemisphere maximum. For the underinformative/true contrast, the larger cluster showed no clear effects over midline central electrode Cz (the depicted waveforms), but the topographic plots show evidence of a frontally-distributed negative-going component. Only the smaller cluster (cluster 2) showed evidence of the reduced N400 effect and extended positivity for the underinformative relative to true items. An important point to note here is that AQ scores and true/false judgments did not differ between any of the clusters, indicating that the variation we see is not clearly a function of autistic-like traits or judgment behavior. Importantly, as with the morphosyntactic and thematic role processing results reported above, these results further underscore the fact that even highly statistically reliable effects seen in grand mean waveforms and analyses do not always depict the normative brain response in the general population under investigation. Moreover, although the variables we investigated (i.e., AQ, truth value judgments) did not modulate these individual differences, using unsupervised clustering algorithms allowed us to observe these neural differences. This opens up a hypothesis space for further work.

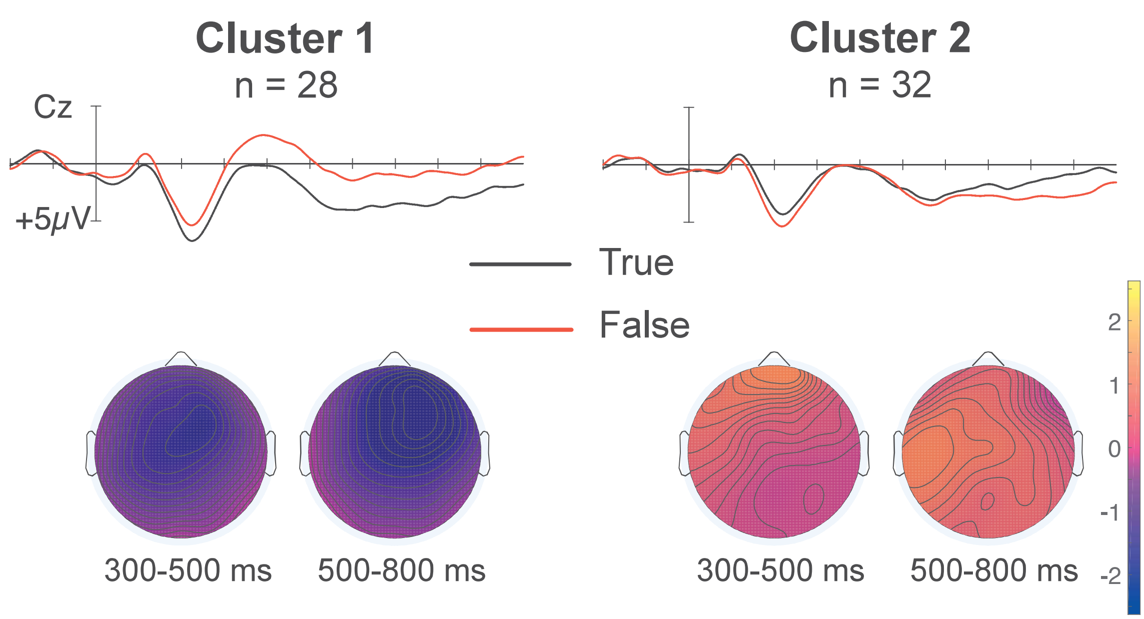

Figure 5. K-means cluster results for the false minus true condition.

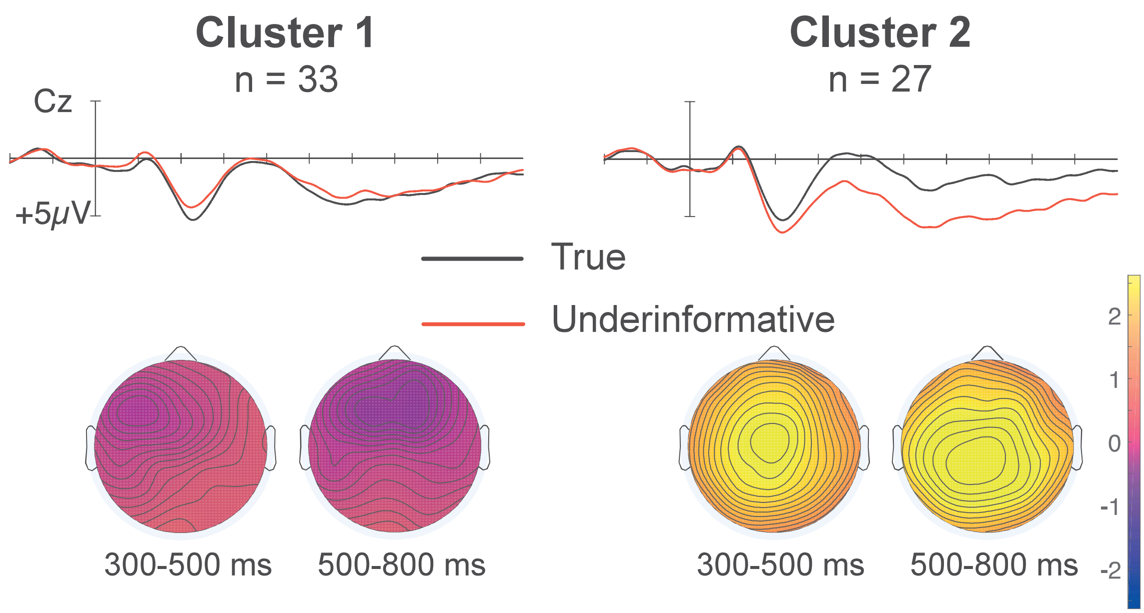

Figure 6. K-means cluster results for the underinformative minus true condition.

Another classic example of communication beyond that which is literally said is irony. Precise definitions of irony have been debated by theoreticians (e.g., Giora, 1997; Kreuz & Glucksberg, 1989; Wilson & Sperber, 1992), but it can be swiftly summarized as when an utterance is used to communicate the opposite of its literal meaning. Uttering *Living with him would be dreadful* may indeed mean living with him is a horrifying thought, but with certain cues (e.g., prosody, situational context, background knowledge, facial expression, emojis), the utterance could be intended to mean that living with him would be wonderful. Heavy reliance on nonliteral cues means irony interpretation is rife with individual differences (Ivanko, Pexman, & Olineck, 2004; LaMarre, Landreville, & Beam, 2009).

One way to indicate irony in text-based communication is the wink emoticon/emoji, linked with irony in both intent and uptake (Filik et al., 2015; Skovholt, Grønning, & Kankaanranta, 2014; Thompson & Filik, 2016). Used in sentence-final position, as in Figure 7C, this wink can serve as a marker of illocutionary force (Dresner & Herring, 2010; Searle, 1969;

Searle, 1979). Here, the wink can mark irony, signaling to the interlocutor that the opposite of the literal meaning is being communicated. Emojis, however, are notably ambiguous (Miller et al., 2016; Miller, Kluver, Thebault-Spieker, Terveen, & Hecht, 2017), and even this conventional ironic wink may not be interpreted as ironic by all individuals nor in all contexts. Interpretation aside, these text with emoji utterances are examples of multimodal communication, in which multiple modalities are used in conjunction in a single utterance (e.g., Cohn, 2016): sentences like those in Figure 7 are not treated as a sentence with an emoji appended to the end, but rather as a systematic multimodal utterance with a single communicative purpose.

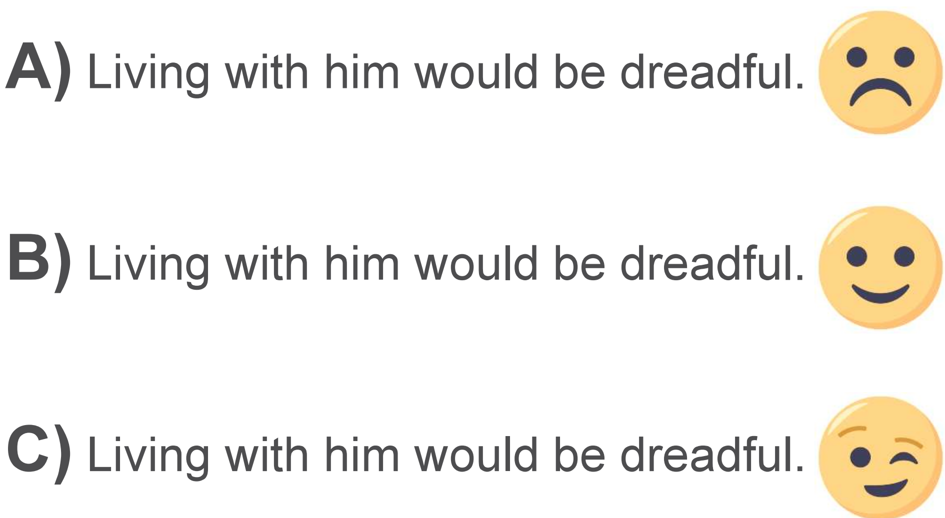

Figure 7. Example sentence followed by a congruent emoji (A), incongruent emoji (B), and ironic (sarcastic) emoji (C).

Previous ERP studies of verbal irony (Filik, Leuthold, Wallington, & Page, 2014; Regel, Coulson, & Gunter, 2010; Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013) used stimuli wherein the final word would be either ironic or non-ironic based on a preceding context. Time-locking to that final word, these studies found a P200 effect and P600 effect to ironic as opposed to literal words (i.e. when that same sentence is preceded by context that leads to a non-ironic interpretation of the sentence). The P200 is typically linked to attention (e.g., Carretie, Mercado, & Tapia, 2001; Luck & Hillyard, 1994), and this P600 finding supports the re-processing/re-analysis theories of the P600 (e.g., Brouwer et al., 2012; Kolk & Chwilla, 2007) as opposed to purely syntactic ones.

We recently completed a three-experiment ERP study (Weissman & Tanner, 2018) in order to study this multimodal way of communicating irony and the individual differences inherent in such a phenomenon. Sentences such as *Living with him would be dreadful* could be followed by either a frown, smile, or wink emoji, representing conditions that either match, mismatch, or render ironic the sentence, respectively (Figure 7). ERPs were time-locked to the sentence-final emoji. After some of the sentences, participants answered a yes/no comprehension question (e.g., *Would living with him be bad?*) designed to determine whether a participant had a literal or nonliteral interpretation of the sentence. If participants say “no” after seeing *Living with him would be dreadful* <*wink*> (i.e., Figure 7C), they are using the wink as a marker of irony to override the literal meaning of the sentence; a “yes” response indicates the participant has not integrated the emoji as a marker of irony and instead adopts the literal interpretation of the sentence.

In the grand average of all 35 participants in Experiment 1, there was a significant P200 effect but no P600 effect. However, by investigating the behavioral responses, we saw a clear and relevant split among participants. Most participants (n = 25) almost never gave nonliteral answers and instead answered according to the literal semantic content of the sentence. A smaller subset of participants (n = 10) gave nonliteral answers over half the time, indicating that they understood the wink emoji as a marker of irony and allowed it to override the literal meaning. Treating these two sets as separate groups paints a different picture. Figure 8A and B show the literal and non-literal groups, respectively. While the P200 effect surfaces for both groups (most strongly over frontal sites), the re-processing P600 effect expectedly showed up only in those participants who actually re-processed and reinterpreted the literal meaning of the sentence. This is a finding that was lost in the overall grand average.

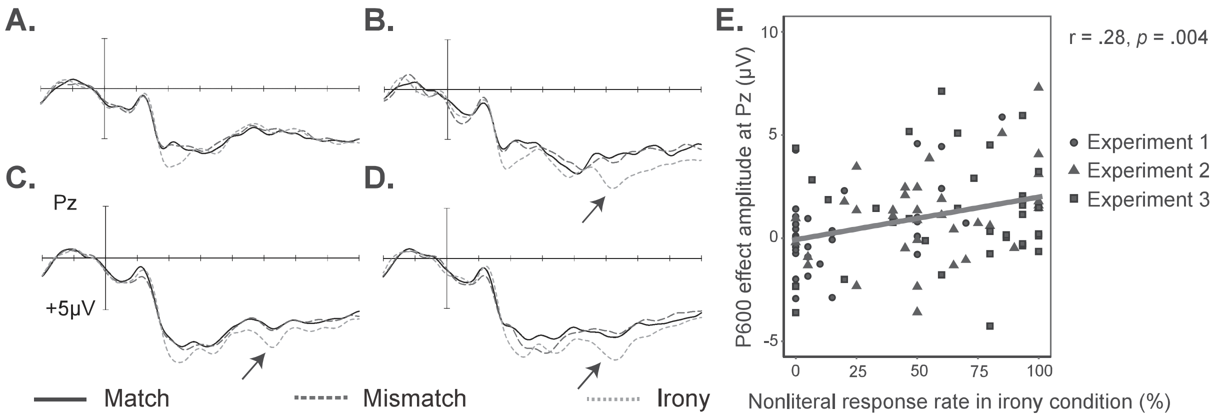

Figure 8. Grand average ERP waveforms at Pz elicited by sentence-final emoji from (A) Experiment 1, literal group (n = 25), (B) Experiment 2, nonliteral group (n = 10), (C) Experiment 2 (N = 35), and (D) Experiment 3 (N = 36). Arrows indicate irony-related P600 effects. (E) Correlation across all three experiments between non-literal response rate and P600 effect amplitude.

Experiment 2 aimed to guide participants towards understanding the sentences nonliterally by mentioning in the instructions that "some of the sentences will be sarcastic." A new set of 35 monolingual English right-handers read the same sentences from Experiment 1 with this additional instruction. The mean nonliteral response rate in irony condition rose from 21.3% in Experiment 1 to 55.0% in Experiment 2. There was no longer a clear split between response groups, and both a P200 and P600 were evident in the grand mean (Figure 8C). Experiment 3 (N = 36) altered the stimulus make-up from Experiments 1 and 2 to rule out a confound in condition-to-emoji mapping, and still found the P200-P600 complex to ironic sentences (Figure 8D).

Figure 8E shows the correlation between nonliteral response rate and P600 effect amplitude at Pz across all three experiments (r = .28, p = .004); across all 106 participants, nonliteral response rate to the yes/no comprehension questions was a significant predictor of P600 effect amplitude. Interestingly, such was not the case for the P200 effect amplitude at Fz (r = .03, p = .76). This investigation into individual differences allows for a better understanding of these components. As only the P600 correlates reliably with comprehension question responses, we can determine that this component is linked to the higher-level processing and interpretation of the stimuli, whereas the P200 effect, present no matter the interpretation, is reflecting a different process (e.g., attention) distinct from these pragmatic interpretations. Unlike morphosyntax, there are not strict rules about the “right” and “wrong” interpretations of irony, emojis, or their intersection. Even so, ERPs have proven to be robustly sensitive to the differing interpretations individuals may have when confronted with potentially-ironic emojis.

Conclusions and Challenges for the Future

As we have demonstrated, although grand mean waveforms and analyses of ERPs can provide meaningful characterizations of the central tendency of brain responses across participants and items, interpreting the grand mean waveform as representative of all (or even most) participants’ brain responses in a study can lead to incorrect conclusions. This is especially the case when strong claims are made about multi-phasic ERP effects in grand means indicating a multi-stage processing architecture in the general population. We have clearly seen that these sorts of conclusions were unwarranted based on biphasic N400-P600 or LAN-P600 responses across a broad range of language user types (novice L2 learners, proficient bilinguals, and literate monolinguals): some participants showed primarily negative-going ERPs, some primarily positive-going ERPs, and some a mix of the two. Thus, even statistical reliability in grand mean analyses does not mean that the observed components generalize to all members of the broader population, and surface component scalp topographies and effect amplitudes can be distorted by component overlap. These findings also extend even to cases where there is a clear monophasic response evident in the grand mean: not all individuals show this response, with some even showing a response of the opposite polarity. Thus, focusing exclusively on parametric analyses of central tendency in ERP measures may lead to inappropriate conclusions about processing in the general population. As we have shown here, some of the 'noise' in ERP measures can indeed be structured and interesting; acknowledging this variation and incorporating it into analyses can lead to more robust and appropriate conclusions. This is similar to how some behavioral researchers are beginning to glean important information about processing by examining not simply mean reaction times, but also distributional properties of reaction times (i.e., ex-Gaussian distributional analyses, e.g., Balota & Yap, 2011; Payne & Federmeier, 2017; Staub, 2010).

A clear consequence of these observations is that extreme care must be taken when interpreting ERP effects in populations that might be expected to differ in some substantive way from proficient, literate, right-handed young adult monolinguals (e.g., children, individuals with language impairments, non-native speakers, bilinguals, older adults, etc). For example, in the field of L2 psycholinguistics, ERPs are often used to identify whether or not L2 processing can be "native-like," where native-like-ness is defined as finding qualitatively similar ERPs in the L2 population as seen in grand mean ERPs for native speakers; any deviations from this are frequently interpreted as indicating aberrant processing in some way, or as evidence of critical period effects in language learning (e.g., N400 effects elicited by morphosyntactic anomalies, or a lack of LAN effects). However, this sort of conclusion is only warranted if no native speakers show this sort of “non-native-like” ERP response. Since we have seen that grand mean ERPs can be problematic to interpret in this way, careful consideration of the range and types of variability in the control population need to be considered. Even for young adult monolinguals, we are just beginning to understand these issues, so broad claims about processing in other populations may be premature. Moreover, the absence of LAN-like components in some populations (like L2 users and bilinguals) cannot be interpreted as indicating the lack of some crucial process, since, as we have seen, many young adult monolinguals do not show LAN effects, and LAN-like components in grand mean analyses can be artifacts of individual differences and component overlap.

The approach we have taken to individual differences in the studies we reviewed here is clearly only one way to approach the problem and begin to understand how variation manifests in language-related ERPs. Clearly there is more variation in the data than our analyses capture, and likely in more dimensions than just the N400/P600 continuum we have highlighted. A challenge for the future, then, is to develop more powerful ways to capture variation in all of the dimensions that ERPs have to offer (e.g., effect timing, polarity, topography, and single-trial dynamics). Note that component overlap makes this necessarily difficult. Here we have documented ways in which components overlap across individuals (e.g., P600s in some cancel out part or all of an N400 in others), but the same issue applies for ERPs computed within individuals, and even electrophysiological effects elicited in single trials: the surface topography of an effect will always reflect the spatiotemporal summation of all of the simultaneously active neural sources of the electrophysiological signal. Because of this, simultaneously active sources within individuals (and trials) can make the nature of the underlying latent components difficult to identify (e.g., even if a LAN effect is found within a single individual or a single trial, it will be difficult or impossible to distinguish it from an N400 effect if there is a simultaneous positivity that obscures part of the negativity, even if that positivity is not clearly evident in the corresponding ERP). This is even the case with blind source separation techniques like independent components analysis, which can misallocate variance for spatiotemporally correlated components (Makeig, Jung, Ghahremani, & Sejnowski, 2000).

Nonetheless, we feel that the studies we have described here provide an important starting point for understanding neurocognitive variation for a range of language phenomena and in a number of populations. We are optimistic that future research on individual differences in language processing will help us better understand the mechanisms and dynamics underlying successful language comprehension, learning, and use very broadly.

References

Antoniou, K., Cummins, C., & Katsos, N. (2016). Why only some adults reject under-informative utterances. *Journal of Pragmatics*, *99*, 78–95. https://doi.org/10.1016/j.pragma.2016.05.001

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, *20*, 160–166. https://doi.org/10.1177/0963721411408885

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (pp. 3–20). New York: Oxford University Press. https://doi.org/10.1300/J108v05n03

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism

Spectrum Quotient: Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. https://doi.org/10.1023/A:1005653411471

Bastiaansen, M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 31–50). New York: Oxford University Press.

Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience*, *33*, 8528–8533. https://doi.org/10.1523/JNEUROSCI.0618-13.2013

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49. https://doi.org/10.1111/j.1467-9922.2006.00353.x

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*, 235–249.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*, 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437–457. https://doi.org/10.1016/j.jml.2004.05.006

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Frontiers in Psychology*, *4*, 1–16. https://doi.org/10.3389/fpsyg.2013.00060

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Carretie, L., Mercado, F., & Tapia, M. (2001). Emotion, attention, and the 'negativity bias', studied through event-related potentials. *International Journal of Psychophysiology*, 75–85.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cohn, N. (2016). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition*, *146*, 304–323. https://doi.org/10.1016/j.cognition.2015.10.007

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*, 172–201. https://doi.org/10.1111/cogs.12227

Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, *20*, 249–268. https://doi.org/10.1111/j.1468-2885.2010.01362.x

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 811–828. https://doi.org/10.1037/a0035658

Filik, R., Turcan, A., Thompson, D., Harvey, N., Davies, H., & Turner, A. (2015). Sarcasm and emoticons: Comprehension and emotional impact. *Quarterly Journal of Experimental Psychology*, *0218*, 1–17. https://doi.org/10.1080/17470218.2015.1106566

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*, 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Grey, S., Tanner, D., & Van Hell, J. G. (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. *Brain Research*, *1669*, 27–43.

https://doi.org/10.1016/j.brainres.2017.05.024

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55. https://doi.org/10.1016/j.cognition.2010.03.014

Hagoort, P., Baggio, G., & Willems, R. (2009). "Semantic Unification." In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (4th Editio, pp. 819–836). Cambridge, MA: MIT Press.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–484.

Hancock, R., & Bever, T. G. (2013). Genetic factors and normal variation in the organization of language. *Biolinguistics*, *7*, 75–95.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*, 101–119.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129–154. https://doi.org/10.1007/BF02172093

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59–73. https://doi.org/10.1016/j.cogbrainres.2003.10.022

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*, 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Ivanko, S., Pexman, P. M., & Olineck, K. (2004). How sarcastic are you?: Individual differences and verbal irony. *Journal of Language and Social Psychology*, *23*, 244–271. https://doi.org/10.1177/0261927X04266809

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: Linguistic processing in high-functioning adults with autism or Asperger syndrome: is local coherence impaired? *Cognition*, *71*, 149–185. https://doi.org/10.1016/S0010-0277(99)00022-0

Jolliffe, T., & Baron-Cohen, S. (2000). Linguistic processing in high-functioning adults with autism or Asperger's syndrome. Is global coherence impaired? *Psychological Medicine*, *30*, 1169–1187.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.

Kim, A., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*, 406–420. https://doi.org/10.1037/xlm0000457

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225. https://doi.org/10.1016/j.jml.2004.10.002

King, J. W., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 376–

395.

Kolk, H. H. J., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, *100*, 257–261. https://doi.org/10.1016/j.bandl.2006.07.006

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, *118*, 374–386.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*, 489–530. https://doi.org/10.1080/01690960500094279

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*, 223–237. https://doi.org/10.1016/j.bandl.2005.12.006

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129. https://doi.org/10.1016/S0926-6410(03)00086-7

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

LaMarre, H. L., Landreville, K. D., & Beam, M. A. (2009). The irony of satire. *The International Journal of Press/Politics*, *14*, 212–231. https://doi.org/10.1177/1940161208330904

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*, 920–933. https://doi.org/10.1038/nrn2532

Lau, W. Y.-P. P., Gau, S. S.-F. F., Chiu, Y.-N. N., Wu, Y.-Y. Y., Chou, W.-J. J., Liu, S.-K. K., & Chou, M.-C. C. (2013). Psychometric properties of the Chinese version of the Autism Spectrum Quotient (AQ). *Research in Developmental Disabilities*, *34*, 294–305. https://doi.org/10.1016/j.ridd.2012.08.005

Lee, C., & Federmeier, K. D. (2015). It's all in the family: Brain asymmetry and syntactic processing of word class. *Psychological Science,* 26, 997-1005. https://doi.org/10.1177/0956797615575743

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.1994.tb02218.x

Makeig, S., Jung, T.-P., Ghahremani, D., & Sejnowski, T. J. (2000). Independent component analysis of simulated ERP data. In T. Nakada (Ed.), *Integrated Human Brain Science* (pp. 123–146). Amsterdam: Elsevier Science.

Martin, I., & Mcdonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *34*, 311–328. https://doi.org/10.1023/B:JADD.0000029553.52889.15

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*, 123–150. https://doi.org/10.1111/j.1467-9922.2010.00604.x

Mehravari, A. S., Emmorey, K., Prat, C. S., Klarman, L., & Osterhout, L. (2017). Brain-based individual difference measures of reading skill in deaf and hearing adults. *Neuropsychologia*, *101*, 153–168. https://doi.org/10.1016/j.neuropsychologia.2017.05.004

Meulman, N., Wieling, M., Sprenger, S. A., Stowe, L. A., & Schmid, M. S. (2015). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. *PLoS ONE*, *10*, e0143328. https://doi.org/10.1371/journal.pone.0143328

Miller, H., Kluver, D., Thebault-Spieker, J., Terveen, L., & Hecht, B. (2017). Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, 152–161.

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). “Blissfully happy” or “ready to fight”: Varying interpretations of emoji. *International AAAI Conference on Web and Social Media*, 259–268. https://doi.org/10.1089/cyber.2011.0179

Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement: A Reply to Tanner et al. *Cortex*, *66*, 156–159. https://doi.org/10.1016/j.cortex.2014.06.009

Molinaro, N., Barber, H., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*, 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Morgan-Short, K. (2014). Electrophysiological approaches to understanding second language acquisition: A field reaching its potential. *Annual Review of Applied Linguistics*, *34*, 15–36. https://doi.org/10.1017/S026719051400004X

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*, 2886–2898. https://doi.org/10.1162/jocn.2009.21400

Neville, H. J., Nicol, J., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*, 151–165.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*, 324–346. https://doi.org/10.1016/j.jml.2010.06.005

Noveck, I. A., & Posada, A. (2003). Charactrerizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*, 203–210.

Noveck, I., & Sperber, D. (Eds.). (2004). *Experimental Pragmatics*. Basingstoke: Palgrave MacMillan. https://doi.org/10.1057/9780230524125

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*, 494–522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and*

Cognitive Processes, *8*, 413–437.

Osterhout, L., McLaughlin, J., Kim, A., Greewald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking-ERPs, and beyond* (pp. 271–308). New York: Psychology Press.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199–230.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkänen, I., … Herschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, *21*, 509–521.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*, 426–448. https://doi.org/10.1016/j.jml.2012.07.003

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*, 2728–2744.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: {E}xperiments at the semantics-pragmatics interface. *Cognition*, *86*, 253–282.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive*

Neuroscience, *10*, 717–33.

Payne, B. R., & Federmeier, K. D. (2017). Pace yourself: Intraindividual variability in context use revealed by self-paced event-related brain potentials. *Journal of Cognitive Neuroscience*, *25*, 837–854. https://doi.org/10.1162/jocn_a_01090

Payne, B. R., Ng, S., Shantz, K., & Federmeier, K. D. (in press). Event-related brain potentials in language processing: The N's and the P's. In A. B. Cieślicka & R. R. Heredia (Eds.), *The Bilingual Brain Unwrapped*. Psychology Press.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383. https://doi.org/10.1080/10888430701530730

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407–428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, *1311*, 121–135. https://doi.org/10.1016/j.brainres.2009.10.077

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*, 277–293. https://doi.org/10.1162/jocn.2010.21411

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press. https://doi.org/10.2307/2184707

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. a, & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*, 2037–2057. https://doi.org/10.1162/jocn.2008.20143

Skovholt, K., Grønning, A., & Kankaanranta, A. (2014). The communicative functions of emoticons in workplace e-mails::-). *Journal of Computer-Mediated Communication*, *19*, 780–797. https://doi.org/10.1111/jcc4.12063

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Spotorno, N., Cheylus, A., Van Der Henst, J. B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS ONE*, *8*, 1–13. https://doi.org/10.1371/journal.pone.0066839

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, *31*, 817–840. https://doi.org/10.1080/23273798.2016.1161806

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, *114*, 447–454.

Steinhauer, K. (2014). Event-related Potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*. https://doi.org/10.1093/applin/amu028

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 397–439). Oxford: Oxford University Press.

Tanner, D. (2018). Robust neurocognitive individual differences in morphosyntactic processing:

A latent variable approach. Manuscript submitted for publication. https://doi.org/10.17605/OSF.IO/NJGT6

Tanner, D., Bulkes, N. Z., Shantz, K., Armstrong, A., & Reyes, A. (2016). Working memory, not language experience, predicts N400 effects to semantically anomalous words during reading. Poster presented at the Neurobiology of Language Conference, London, UK.

Tanner, D., Grey, S., & Van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*, 248–259. https://doi.org/10.1111/psyp.12788

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*, 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*, 367–382. https://doi.org/10.1017/S1366728912000302

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215. https://doi.org/10.1016/j.jml.2014.07.003

Tanner, D., & Van Hell, J. G. (2014a). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002

Tanner, D., & Van Hell, J. G. (2014b). Gradience and individual differences in processing syntactic, semantic, and thematic information: ERP evidence. Paper presented at the 2014 CUNY Conference on Human Sentence Processing, Columbus, OH, USA.

Thompson, D., & Filik, R. (2016). Sarcasm in written communication: Emoticons are efficient markers of intention. *Journal of Computer-Mediated Communication*, *21*, 105–120. https://doi.org/10.1111/jcc4.12156

Townsend, D. J., Carrithers, C., & Bever, T. G. (2001). Familial handedness and access to words, meaning, and syntax during sentence comprehension. *Brain and Language*, *78*, 308–331. https://doi.org/10.1006/brln.2001.2469

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*, 124–139. https://doi.org/10.1111/psyp.12299

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*, 67–82.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*, 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*, 1181–1197. https://doi.org/10.1162/jocn.2006.18.7.1181

Weissman, B., & Tanner, D. (2018). A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension. *PLoS ONE*.

Wilson, D., & Sperber, D. (1992). On verbal irony. *Lingua*, *87*, 53–76.